DOCUMENT RESUME

ED 404 354 TM 026 144

AUTHOR	Devine, Marion; And Others
TITLE	Maintaining Standards: Performance at Higher Grade in Biology, English, Geography & Mathematics. SCRE Research Report No. 78.
INSTITUTION	Scottish Council for Research in Education, Edinburgh.
SPONS AGENCY	Scottish Office Education and Industry Dept., Edinburgh.
REPORT NO	ISBN-1-86003-028-9
PUB DATE	96
NOTE	131p.; For a summary report, see TM 026 156.
AVAILABLE FROM	Scottish Council for Research in Education, 15 St. John Street, Edinburgh EH8 8JR, Scotland, United Kingdom; e-mail: SCRE@ed.ac.uk (10 British pounds).
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement; Achievement Gains; Achievement Tests; *Biology; Change; *English; Foreign Countries; *Geography; *Mathematics; Secondary Education; Secondary School Students; *Standards
IDENTIFIERS	*Scotland; Subject Content Knowledge

ABSTRACT

In June 1994 the Scottish Office Education Department asked the Scottish Council for Research in Education to assist in a study of performance over time in the Higher Grade examinations in biology, English, geography, and mathematics. Statistics have shown that the number of students gaining passes on these Higher Grade examinations is increasing from year to year, and the question of whether the standards for these examinations have been eroded has been raised. The study depended on the analysis of original test scripts retained by the Scottish Examination Board, beginning with 1987 as the earliest year for which scripts were available for all subjects. In general, 100 scripts were used for each subject for each year; variations that caused small reductions in the actual number of scripts are noted in the sections describing each subject. The analysis supports the conclusion that there are no grounds for believing that there has been any change in the standards of performance for any given band of award (ranging from A to no award). Some differences in detail for specific subjects are noted in sections reviewing the scripts by content area. An appendix for each of the subjects describes results in detail. (Contains 14 figures, 31 tables, 9 appendix tables, and 146 appendix graphs.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 404 354

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
E. WAKE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Maintaining Standards

Performance at Higher Grade in Biology, English, Geography & Mathematics

Marion Devine
John Hall
Jacqueline Mapp
Kerry Musselbrook

TM026144

BEST COPY AVAILABLE

Maintaining Standards

Performance at Higher Grade in Biology, English, Geography & Mathematics

Marion Devine
John Hall
Jacqueline Mapp
Kerry Musselbrook

SCRE
The Scottish Council for
Research in Education

SCRE Research Report No 78

ISBN 1 86003 028 9

Copyright © 1996 The Scottish Council for Research in Education

First published October 1996

A report arising from the research project 'Performance in Higher Grade Examinations' (Ref: R/13/3) funded by the Scottish Office Education and Industry Department and the Scottish Council for Research in Education from June 1994 to June 1996.

The views expressed are those of the authors and are not necessarily those of the Scottish Council for Research in Education or the Scottish Office Education and Industry Department.

Printed and bound by the Scottish Council for Research in Education, 15 St John Street, Edinburgh EH8 8JR.

Contents

Page

Acknowledgements vii

Findings and observations viii
The questions viii
The answers viii
  Biology ix
  English ix
  Geography ix
  Mathematics ix
Observations on the exercise x

1. Introduction 1
The context for the study 1
Aims of research 1
Issues affecting all subjects 1
  Differences between revised and unrevised examinations 1
  The identification of criteria for analysis 2
  Differences between this project and SEB practice 3
Stages of the project 3
  Consultation with SOEID and SEB 3
  Analytic frameworks 4
  Training of markers 5
Subject areas 5

2. Biology 6
Structure of the syllabus and examination 6
Essays 8
Limitations on the study 9
Procedure 11
Findings 11
The content areas 12
The essay questions 14
Conclusions 16
Appendices 17
  B.1: Marking schemes for essays 18
  B:2 Distribution of scores in each skill area for each band of award 20

3. English 24
Methodology 25
  Stage 1: Drafting grading schedules 25
  Stage 2: The re-marking of scripts 26
The sample 26
Trial marking 26
The component parts 27
The findings 29
  Individual criteria 29
  Language skills 36
Conclusions 37
Appendices 38
  E.l: Marking grids used in the research 39
  E.2: Criteria for categorising Interpretation papers 43
  E.3: List of criteria compared across years together with results of statistical tests 45

4. Geography 53
Differences between unrevised and revised Higher Grade geography 53
Sample size and type 54
Framework for analysis 54
Levels of attainment 56
The marking process 56
Analysis of results 57
  Scatter of attainment 57
  Different forms of analysis 58
  Between year analysis 60
Summary 62
  Methodological issues 62
  Differences in performance between candidates at bands A to D 62
  Range of performance 63
  Differences between the years 63
Conclusions 63
  Recommendations 64
Appendices 65
  G.1 Cumulative profiles illustrating the differences between the bands 66
  G.2 Cumulative profiles illustrating the differences between the years 73

5. Mathematics **80**
Differences between unrevised and revised Higher Grade mathematics 80
  Content 80
  Approach 81
  Format of examination 82
Basis for comparison 82
Sample size and type 82
The marking process 83
Analysis of results 84
  Attempt rates 84
  Attainment on non-comparable criteria 85
  Attainment on individual comparable criteria 85
  Attainment on basic grouped criteria 88
  Attainment on groups comprising all instances of relevant criteria 91
  Presentation and the ability to work logically 95
Summary 96
  Contrasting unrevised and revised examinations 96
  Discrete criteria 96
  Grouped criteria based on straightforward instances of criteria 97
  Grouped criteria based on all instances of criteria 97
Conclusions 98
Appendices 99
  M.1 List of all criteria used in the analysis of the scripts 100
  M.2 Bar graphs showing comparisons between the performance of candidates in unrevised and revised Higher Grade examinations on individual criteria 104

List of Tables

1.1 Number of candidates' scripts available for analysis 4
2.1 Question categories in the Higher biology examination papers 7
2.2 Comparable essay questions on industrial melanism 8
2.3 Comparable essay questions on photosynthesis 8
2.4 Higher Grade biology sample 9
2.5 Performance in Higher Grade biology skill areas 12
2.6 Number of knowledge and understanding items identified for each content area in the examination paper for each year 13
2.7 Photosynthesis essay, 1988 and 1994 14
2.8 Industrial melanism essay, 1987 and 1994 14
3.1 Higher Grade English: Sample of scripts 26
3.2 Higher Grade English: Interpretation 34
3.3 Total number of items addressing each criterion in the Interpretation papers 35

List of Tables (contd)

4.1 Unrevised and Revised Higher Grade examinations 53
4.2 Geography sample 54
4.3 Opportunities offered in the different papers 56
4.4 Indicator 2 - the ability to explain 57
4.5 Converting mean scores to attainment levels 58
4.6 Percentage of candidates demonstrating attainment in the ability to describe 59
4.7 Band A candidates' ability to describe 60
4.8 Band A candidates' ability to use graphic skills 61
4.9 Significant differences between the years 61
5.1 Sample of mathematics scripts by band 83
5.2 Percentage of candidates demonstrating attainment 85
5.3 List of comparable criteria 87
5.4 Number of differences 88
5.5 Opportunities and percentage attempt rates for grouped criteria 92
5.6 Attainment when non attempts are included as failure 93
5.7 Correct responses as a percentage of those criteria attempted 94
5.8 Significant differences 95
5.9 Percentage of candidates able to set out work logically 96
5.10 Percentage of candidates able to present work neatly 96

List of Figures

2.1 Distribution of band awards for Higher Grade biology 1987-1994 10
2.2 Distribution of band awards within the sample available to the project 10
2.3 Performance of band C candidates on photosynthesis essay 15
2.4 Performance of band A candidates on photosynthesis essay 15
3.1 Discursive writing (including personal reflection) - examining language skills 30
3.2 Imaginative Writing - band A - language criterion (SD) 32
3.3 Imaginative writing - band A - tone/mood criterion (SD) 33
3.4 Language skills - Band C (NSD) 37
4.1 Cumulative frequency graphs for the indicator 'describe' 59
4.2 Band A candidates' ability to describe 60
5.1 Basic algebra manipulations 89
5.2 Basic geometry 90
5.3 Basic trigonometry 90
5.4 Basic calculus 91

Acknowledgements

We would like to thank all the teachers and the subject specialists from the Inspectorate and the Scottish Examination Board who provided their time and insights during the course of the study. We particularly appreciate the high level of co-operation from the SEB without which the project could not have taken place. Four Examination Officers of the SEB supported the project with their expertise throughout the study.

Special thanks are also due to Anne Brand and her staff from the SEB for the prompt and efficient way in which they collated the sample of scripts from the archives.

We would also like to acknowledge the hard work of our project secretary Margaret Caddick in preparing the final version of the report.

Finally we would like to express our appreciation to the markers for their hard work and conscientiousness in helping to refine and then implement the different marking schemes. The complete list of markers for each subject is noted below. Those marked with an asterisk also acted as subject consultants throughout the project.

Marion Devine
John Hall
Jacqueline Mapp
Kerry Musselbrook

August 1996

Findings and observations on performance in Higher Grade examinations

The questions

This research set out to answer two basic questions about performance in Higher Grade examinations. These were:

> Have there been any significant changes in the standards of attainment required to gain a pass in Higher Grade examinations in biology, English, geography and mathematics in recent years?

And, if there were any identifiable variations in standards:

> Can we identify any factors responsible for these variations?

The research project had available to it a sample of candidates' scripts in each subject from four years chosen between 1987 and 1994. This sample forms part of the archive of candidates' scripts from past years kept by the Scottish Examination Board. These years straddled the introduction of the revised Higher Grade examination in each subject. Methodological details of the project are in the body of the report.

The answers

The overall conclusion to be drawn from the studies presented in this report must be that:

- from the evidence available to this research project there are no grounds for believing that there has been any change in the standards of performance required to obtain any given band of award in the Higher Grade examinations which formed part of the study.

This is the overall conclusion. In matters of detail there were some differences in performance between years in particular subjects which did become apparent. Further investigation of these, however, suggested that they could be ascribed to:

- particular features of the examination papers in particular years
- difficulties caused by relatively small sample sizes for particular aspects of the study
- changes to syllabuses which meant that like was not being compared with like
- individual question differences.

The findings relative to each subject are given below. These are reproduced here from the conclusions in each subject chapter, to which the reader is referred for further detail. A general point to bear in mind is that there are two types of finding: the first are judgements about the overall standard of performance in the examination evident from the data available; the second are findings relating to particular parts of the examination (eg a paper), particular criteria assessed within the examination, or particular aspects of the syllabus covered by the examination.

Biology (chapter two)

1. There was no evidence to support the view that there has been an overall change in standards in Higher Grade biology in the years covered by the study. Some statistically significant differences were observed in performance between years in particular skill areas at particular bands of award. However these did not tend in any one direction or indicate any general shift in standards.
2. The more detailed study afforded by the specific examples of the two repeated essays (see chapter two) supports the view that there has remained a consistency of standards at almost all bands of award.
3. There are some suggestions, though they are not conclusive, that the very best candidates are exhibiting higher levels of performance in the later years of the sample.

English (chapter three)

1. There is no evidence for any overall variation in standards in the years covered by our sample in Higher Grade English. The performance represented by each band of award appears to remain consistent.
2. The very few differences in performance on individual criteria which were found were not indicative of any general pattern and were likely to be explained by features of the sample or the particular papers set in each year.

Geography (chapter four)

1. There was no evidence to suggest that standards of pupil performance between the unrevised and revised examinations had changed.
2. There were no significant differences between years in the ability of candidates to discuss or in their ability to demonstrate understanding of geographic processes.
3. Where there were significant differences, these tended to be between 1987 and 1991 *[both unrevised examinations]* rather than between unrevised and revised examinations.
4. When significant differences were studied in relation to separate papers all differences were in Paper I.

Mathematics (chapter five)

1. At the level of individual criteria, which were judged to be directly comparable between the unrevised and the revised examinations, no significant differences were noted.
2. When comparisons were made in relation to straightforward instances of criteria grouped under the headings of algebra, straight line geometry, trigonometry and calculus, there were no significant differences.
3. When all instances of relevant criteria were included and comparisons were made between performances in the revised and unrevised examinations, there were significant differences at bands A, B and C in integral calculus and in bands B and D in differential calculus. In each case

there were more correct responses in the unrevised examinations. Possible explanations for these differences are discussed in chapter five.

4. When all instances of relevant criteria were included and comparisons were made between performances in the revised and unrevised examinations, there were significant differences at all bands in geometry. In each case there were more correct responses in the revised examinations.
5. There were no significant differences in candidates' ability to present their work clearly or set out their work in a logical manner.

■ Observations on the exercise

Throughout this project we were encouraged to regard it as, at least in part, a feasibility study for similar exercises in the future. To that end, we were also asked to reflect upon our experience in conducting the research and to make any suggestions we thought appropriate for such future work. We start by observing that, while the Scottish Examination Board has meticulously archived samples of candidates' examination scripts for many years, this is the first occasion that we, or they, know of where this archive has been used. The Board conducts exhaustive statistical analyses of examination results and candidates' performance on individual papers and items within the examinations, but the historical archive has not previously been used for any study of performance over time. The Board themselves have been interested to learn whether the archive has been useful. We can assure them that it has and thank them for their support for, and interest in, the research.

Any suggestions or recommendations must begin by noting that there is a need to ask about the purpose of the archive. Why is it being kept? Is it an illustrative record of typical performance? If so, what is meant by 'typical'? What should be the relationship between the sample and the population from which it is drawn? Is it meant as a resource for research, a check on standards, an internal reference for the Board? The current research has made only one out of a range of possible uses for the archive of scripts. To take only one very particular example of an alternative use which occurred to the research team, it would have been fascinating to have studied the range of texts chosen by candidates in Higher Grade English, and checked to see how these had changed (if they had). We were not, of course, asked to study that particular question, but the general point is that whatever purpose is envisaged for the archive may well affect what is kept, and how it is kept. It is also possible that designing the archive to fulfil only one specialist purpose might detract from its usefulness for other purposes. All of these questions need to be addressed systematically and explicitly.

There is also a need to think about the structure of the sample of scripts which is kept. At present the Board sample randomly within each range mark (several of which make up a band of award). Since equal numbers of scripts are kept within each range mark, and the spread of range marks within each band of award varies, there is an unusual distribution of scripts between bands of award (see the illustration of this in chapter two). Since the band of award was a major unit of analysis in the current research, this particular distribution was not ideal for our purpose. It also seems to us that the number of scripts kept in each range mark in any one year is too small to form a meaningful sample for statistical analysis (indeed, there were occasional problems with the numbers in the sample for particular bands of award). It might be, therefore that the sample should either be structured around a unit other than the range marks, or that the sample size should be increased.

A further question arises about whether the sample should be constructed around range marks (or band awards) at all, or whether it should reflect the distribution of performances in the population as a whole. This would certainly be beneficial to any study which was only concerned with the mean performance, or with the boundary between 'pass' and 'fail'. However it would probably require a very much larger sample size than at present if any meaningful conclusions about the extremes of performance were to be reached. This may be impractical. There is therefore a need to think closely about the criteria for selecting scripts for the sample.

There was some data which was not available to us. A problem which was particularly evident in the study of performance in Higher Grade biology was the lack of individual candidates' responses from Paper I in that examination. Paper I (biology) is a multiple choice item-banked paper in which the items are drawn from a confidential bank of items, many of which are re-used from year to year. This is a device to enhance consistency within the examination between years and is not, we understand, confined to biology. We would suggest that it may be possible to associate candidates' results in such papers with their Paper II scripts in the archive, thus making a wider range of data available to future researchers. It may even be possible to archive copies of answer sheets from such papers with the scripts. This is a question for the Board to consider.

Another feature which was suggested by our experience with the Higher Grade biology study was the usefulness (to this exercise) of repeated essay questions. The particular questions which were repeated in the years we studied were not exactly identical with one another but were close enough for us to make valid comparisons. If the archive is to be used in the future for checking standards over time, then the usefulness of these repeated questions should be noted.

The research was unable to take any account of internally assessed elements of the Higher Grade. We therefore have no way of knowing to what extent this contributed to the final awards of the candidates. We confess that we can think of no simple solution to this, but the existence of internally assessed elements should be remembered and considered in any future exercise.

Similarly, future researchers should give very careful thought to the type and scale of the criteria which they use to analyse candidates' performances. In these studies we have been guided by the panels of subject experts who worked closely with us. However, this has resulted in different approaches being taken in different subjects. We have examples of large numbers of very finely-focused small-scale criteria, and other instances of large, relatively general criteria. There are dichotomous criteria (marked as a simple 'pass/fail') and others which are expressed in terms of graded levels of performance. This very much means that each subject is a separate study. In the choice between uniformity of approach and reflecting the nature of the subject (and examination) as understood by practitioners, we opted for the latter. It may be that future researchers will disagree. There is, however, a choice to be made and much thought that needs to be given to the reasons for that choice.

If it were decided in the future that there should be more comparability of approach between different subject areas then we would most certainly recommend that a large amount of preparatory time should be set aside for such a purpose. The development of uniform methods, applicable to a range of subjects, and yet capable of being judged as appropriate to the subject by subject experts, would be no easy task. It would require substantial input from the researchers and the subject experts, and extensive liaison between experts in different subject areas to reach agreement about possible common methods.

With the changes now being made in relation to *Higher Still* careful thought should be given to the possibility of undertaking a similar exercise before the new structure comes into place. We now have a base-line of information on the revised examinations and it would be a pity to lose the opportunity of following the results through with the present system.

1 Introduction

In June 1994, the Scottish Office Education Department asked the Scottish Council for Research in Education to collaborate with members of HMI in the study of performance over time in Higher Grade Examinations in four subjects - biology, English, geography and mathematics. The report of the study is in two main sections. The first section deals with matters common to all four subjects, the second provides detailed information on each separate subject.

■ The context for the study

There is a continuing interest in the attainment of pupils and whether this is changing over time. For the primary and early secondary years, the Assessment of Achievement Programme provides regular monitoring information about Primary 4, Primary 7 and Secondary 2 pupils. Higher Grade examination results provide for many pupils the ultimate measures of their attainments and this study aims to consider the extent to which standards of performance are being maintained.

Statistics show that the numbers of pupils gaining Higher Grade passes are increasing from year to year. The Scottish Examination Board (SEB) has set up procedures and trained staff to ensure that Higher Grade questions reflect the syllabus and that pupils' answers are marked to a consistent standard between candidates and from year to year. However, changes to the syllabus and to the format of questions may make it difficult to achieve high levels of consistency of pupil performance over a period of several years.

■ Aims of research

The main aim of the study was to

- determine if there have been any significant changes in the standard of attainment required to gain a pass in Higher Grade examinations in biology, English, geography and mathematics.

If there were found to be variations in standards,

- the research will attempt to identify the factors responsible.

■ Issues affecting all subjects

Several of the issues which were the concerns of the four subject working parties were common to all. These included differences between unrevised and revised examinations, the identification of criteria for pupil performance which could be used across time and the differences between the analysis for this project compared with the way in which the SEB makes its awarding decisions.

Differences between revised and unrevised examinations

Consistency of content and approach between unrevised and revised Higher examinations differed from subject to subject. In English, the core elements had remained relatively consistent, making it possible to draw direct comparisons on a wide range of criteria over the years from which the sample was drawn. In biology the differences were considered to be mainly in terms of presentation and emphasis with a substantial amount of commonality between the two forms of the examination.

The differences between the two examinations in geography were also more to do with emphasis than a radical change of content but the change in format meant that candidates had less choice in the revised than the unrevised examination.

The changes in the mathematics examinations were considered to be much more radical. The differences occurred in the content of the syllabus, in the approach to mathematics as well as in the format of the examination. The basis for direct comparison was more restricted for mathematics than for the other three subjects.

The greater the difference between the examinations the more difficult it was to select criteria which could be used to analyse pupil performance over the years.

The identification of criteria for analysis

At an early stage in the project one of the topics for discussion by all subject working groups was to decide about the scale of the criteria to be used for the analysis. A decision had to be made as to whether the criteria ought to be sufficiently detailed to allow a comparison between performance on specific knowledge and skills or whether a more broad brush approach should be taken to allow the analysis of performance on fundamental principles of the subject. It was decided that, as this was the first time any such longitudinal analysis had been conducted, each subject would choose its own type of criteria and that there would be no need to impose uniformity across the subjects.

In English, because of the consistency of approach over the years, it was relatively easy to adapt the existing framework of the revised Higher Grade course for use across the years. These criteria consisted of grade related statements of performance on key components of Higher Grade English.

In biology, it was decided that a number of broad skill areas - knowledge and understanding, problem solving and practical applications - were common to both the unrevised and the revised Higher Grade examinations as was the basic content. It was decided therefore to analyse performance using both skills and content.

The working group in geography decided against the analysis of content as this changed from year to year but opted instead for a small number of indicators of ability which are fundamental to the study of geography. Seven such indicators were identified and are reported in the relevant chapter.

In mathematics, early scrutiny of the papers showed that it was difficult to make comparisons over the years on a broad fundamental skill such as 'problem solving' as the approach to mathematics had changed so much. The criteria used to analyse the papers are, therefore, small in scale and provide evidence for detailed comparison of performance. Results from the small-scale criteria have been aggregated to provide evidence for attainment in a number of topic areas. Further information is included in the relevant chapter.

Both the nature of the subject and the extent of change over the years from which the examination scripts were drawn played a part in determining the kinds of criteria selected.

Differences between this project and SEB practice

For the SEB, decisions about the overall award to candidates depend on an holistic approach. Candidates awarded an A band do not have to perform well on every skill assessed as part of the examination. It is their overall performance which determines their award. In this project, although the overall awards were used as a sorting factor to make comparisons across years and within years, it was performance on specific criteria that was being studied. Some discrepancies could, therefore, be expected. For example, for some individual criteria, there might be no difference between students awarded band A and band B.

The award by the SEB is based on performance on all aspects of the examination. In this project, certain aspects of each examination were not included. In some cases such as Practical Criticism in English and Aspects of economic geography it was because the overall number of candidates selecting these options was small.

In Biology, Paper 1 is a multiple choice paper and although the SEB retains statistics on these questions, individual responses from our sample of candidates were not available. In the unrevised Higher Grade examination in mathematics, Paper 1 was a multiple choice paper. In the revised Higher Grade examination, Paper 1 was an open response paper. Even if individual candidates' responses had been available for the multiple choice paper, the different format would have made it difficult to compare performance.

In mathematics and geography, an investigation is included as part of the Higher Grade assessment. In both cases the investigation was not included in the project's analysis.

■ Stages of the project

Consultation with SOEID and SEB

Members of HMI with appropriate subject expertise advised and supported the research at every stage. The SEB also took an active role in the study by providing the expertise of Examination Officers as well as releasing all necessary scripts and copies of other documents as requested. Working groups were set up consisting of researchers, members of HMI and Examination Officers in each subject to take forward the work of the project.

The study depended on the analysis of original scripts retained by the Scottish Examination Board in its archives. Originally it had been intended to compare scripts over a ten year period from 1983 to 1993, however, consultations with the SEB indicated that 1987 was the earliest year in which a sample of scripts could be obtained for all four subjects.

In the years from 1987 until 1992, 100 scripts had been retained by the SEB in each of the four subjects. In 1993, the same number had been retained for English and mathematics but the sample for biology and geography had been reduced to 50 scripts. It was, therefore, agreed that 1994 scripts should be included in the study and the SEB retained 100 scripts from each of the four subjects for this purpose.

The scripts were selected randomly by SEB within range marks. These range marks are linked to bands as follows:

Range mark	Band
1-5	A
6-7	B
8-9	C
10-11	D
12 - 14	No award

Since five range marks are assigned to band A and only two to B, C and D the sample contained more band A scripts than other bands. Table 1.1 shows the approximate number of candidates' scripts at each band which were available in each subject.

Table 1.1: Number of candidates' scripts available for analysis

Band	*A*	*B*	*C*	*D*	No award
No of scripts	36	16	16	16	16

The actual number of scripts analysed in each subject is provided under the subject specific sections of this report.

It was essential for the study to cover both unrevised and revised Higher Grade examinations. The first revised Higher Grade examination was introduced in 1987 in mathematics but only a small number of candidates sat this examination. By 1993, unrevised Higher Grade examinations were still available in biology and geography but again these were taken by small numbers only. The years from 1987 to 1994 included the revised/unrevised syllabuses for all subjects.

There was also a concern that, in any one year, only a sample of the syllabus can be covered by the examination. It was, therefore, decided to select two consecutive years for study from the unrevised Higher Grade examinations and two from the revised Higher Grade examinations. For most subjects 1987/88 (unrevised) and 1993/94 (revised) were selected. Any deviations from these years are explained in the subject sections.

Analytic frameworks

A major task for all subject groups was to determine the framework which would be used to analyse the scripts. The task was not to re-mark the scripts using the marking schemes of the SEB, which would have differed from year to year, but to re-analyse the papers using a single set of criteria for each subject. Working groups in each subject used the same general principles. It was agreed that the analytic framework

- should be recognisable to subject experts as a working synthesis of the subject as taught
- should comprise criteria which are discrete and observable from the scripts
- should achieve a balance between the need to describe detail and the need to work with fundamental aspects of the subject
- should be sufficiently robust to accommodate changes in the subject over the years covered by the project.

The details of the criteria used for the analyses are provided under the separate subject headings in this report

Training of markers

Each subject set aside one half day or one whole day for the training of markers. In each case, the markers were experienced people who had been involved in marking examination scripts at Higher Grade for several years. The training session allowed them to discuss the criteria to be used and to practice marking on selected scripts. In each case the markers agreed that the criteria used were appropriate for the subject and could be used to analyse performance. Different numbers of markers were used for different subjects and the marking process took about 3/4 weeks. Various proforma and marking grids were developed to make it easy for markers to record their analysis of the scripts and for researchers to transfer the records to computer.

■ Subject areas

The criteria and the procedures used to analyse each subject area are provided under subject headings in the remainder of the report. The main findings and a commentary on these findings are included. Where significant differences are noted, the level of significance is set at $p < 0.01$ ie we can be 99% confident that the differences have not occurred by chance. Graphical and technical details which are provided as appendices are attached to the end of each subject section.

2 Biology

When the biology working group first met there was considerable discussion of the differences between the unrevised and revised Higher Grade examinations. However, these tended to be differences of presentation and emphasis. Upon investigation, it was thought that there was a considerable amount of commonality between the two forms of the examination.

■ Structure of the syllabus and examination

The revised Higher Grade biology examination is intended to assess knowledge and understanding, problem solving and investigation or practical application. The revised Higher is structured in two papers, the first of which is a multiple-choice paper which draws on a bank of questions kept by the SEB. No completed question papers from this multiple-choice paper are held by the SEB. While some statistics on these questions are kept, it was not possible to conduct any useful analysis of these statistics as part of this project principally because it would be impossible to link this data directly to the scripts for our sample of candidates available from the archive.

Paper II of the revised Higher contains a series of short structured questions designed to test knowledge and understanding; longer sets of questions designed to test data handling known as problem solving and some aspects of practical work previously undertaken known as practical applications; and essay questions on a variety of topics. The working party agreed that this broad characterisation of question types into knowledge and understanding, problem solving and practical applications was equally applicable to the earlier, unrevised, version of the Higher Grade examination. We refer to these three categories as the skill areas in this report.

The revised Arrangements for the Higher Grade biology examination lists the major topics included in the syllabus as:

- Cell biology
- Genetics and evolution
- Control of growth and development
- Regulation in biological systems
- Adaptation

Each of these topics is, in turn, divided into a number of sub-topics. There have been some changes in terminology, content and emphasis between the unrevised and revised Higher Grade examinations, but the working party felt that these topics and sub-topics reflected the basic content structure of both unrevised and revised Higher Grade biology and could therefore be used as the basis of an analysis of the papers from all years in the sample. The questions could also be categorised by the skill areas which they were addressing. We were able to confirm that this scheme could be applied to all years and would cover by far the greater part of the content of the examinations. There was, however, some content in some papers which was not covered by the analysis. The grid used for categorising the questions is shown in Table 2.1.

All questions in the relevant papers from 1987, 1988, 1993 and 1994 were categorised using this grid. These categories then formed the basis of subsequent analysis.

Table 2.1: Question categories in the Higher Grade biology examination papers.

Skills: Content: **Topic** sub-topic	Questions relating to **knowledge and understanding** of the topic	Questions relating to **problem solving** or **handling information** in this topic area	Questions relating to **practical** work or **investigations** which have been undertaken in this area
Cell biology			
• Cell variety in relation to function			
• Absorption and secretion of materials			
• Photosynthesis			
• Energy release			
• Synthesis and release of proteins			
• Cellular response in defence			
Genetics and evolution			
• Variation			
• Selection and speciation			
Control of growth and development			
• Growth differences between plants and animals			
• Genetic control			
• Hormonal influences			
• Environmental influences			
Regulation in biological systems			
• Physiological homeostasis			
• Population dynamics			
Adaptation			
• Maintaining a water balance			
• Obtaining food			
• Coping with dangers			

Questions in the examination paper were categorised according to which skill area they addressed, and which content topic (and sub-topic) they covered.

Essays

Also included in the Higher Grade biology examination, in both the unrevised and revised versions, were a number of questions which require short essays in response. These could be (and were) categorised according to the above scheme. However, the papers for the relevant years were found to include two pairs of questions which were on similar topics and demanded similar responses. In 1987 and 1994 there were questions which demanded knowledge of industrial melanism, while in 1988 and 1994 there were questions on photosynthesis. The complete questions are reproduced in Tables 2.2 and 2.3, with the directly comparable sections in italics.

Table 2.2: Comparable essay questions on industrial melanism

1987 Paper II, Section C

Question 13A

Give an account of evolution **under the following headings**:

(a) The theory of Lamarck;

(b) Continental distribution of organisms;

(c) Analogous structures;

(d) Industrial melanism as observed in the peppered moth.

1994, Paper II, Section C

Question 14A

Using named examples, discuss selection under the following headings:

(a) Evolution of resistant strains of organisms;

(b) Industrial melanism;

(c) Development of varieties of organisms by artificial selection.

Table 2.3: Comparable essay questions on photosynthesis

1988, Paper II, Section C

Question 13A

Discuss the light and dark stages of the photosynthetic process. Describe briefly how a green leaf is adapted for this process.

1994, Paper II, Section C

Question 15B

Give an account of the light and dark stages in photosynthesis

Although the questions are not identical, it was felt that they made similar demands of candidates and that they afforded an opportunity for direct comparisons of candidates' performances over the years. Marking schemes were therefore constructed which were equally applicable to each occurrence of the questions. These schemes were constructed by SEB principal examiners, and were used by them to re-mark the candidates' responses (see Appendix B.1). This allowed direct comparison between performance in 1987 and 1994 (for the question on industrial melanism); and 1988 and 1994 (for the question on photosynthesis).

Limitations on the study

There are two major limitations on this study of Higher Grade biology. The first is caused by the structure of the examination and the unavailability to us of data from Paper 1 (the item-banked multiple choice paper). Candidates' performance on Paper 1 will have a direct influence on the final band which they are awarded, and it is these bands which are used in our analysis. The precise contribution of performance on Paper 1 to the final band of award for each candidate in our sample was not known.

The second limitation is caused by the structure of the sample which was available to us. The sample for 1993 was less than half the size of that for other years, as is shown below in Table 2.4.

Table 2.4: Higher Grade biology sample

Year	Band A	Band B	Band C	Band D	No award	Total
1987	33	16	16	16	19	100
1988	36	16	16	15	17	100
1993	15	7	8	8	9	47
1994	34	16	16	16	18	100
Total	118	55	56	55	63	347

This imbalance was caused by the fact that in 1993 the SEB archived 100 scripts, but they were drawn from both the unrevised Higher (which was still running) and the revised Higher. At an early stage in this project it was decided that the latter two years of the sample should draw on only scripts from candidates who had sat the revised Higher. However, this comparatively small sample, especially so when much of the analysis was conducted at the level of the banded awards, must raise severe questions about any findings which involve the 1993 sample.

Another feature of the sample which must be borne in mind is that, because of the way that the SEB archives material, the structure of the sample available to us does not reflect the distribution of band awards of the population as a whole. This is illustrated in Figures 2.1 and 2.2 below. The effect of this on the project was to restrict all analysis to comparisons within bands of award. This feature of the sample applies to all subjects in the current study.

Figure 2.1: Distribution of band awards for Higher Grade biology 1987 -1994. (Source: SEB Reports)

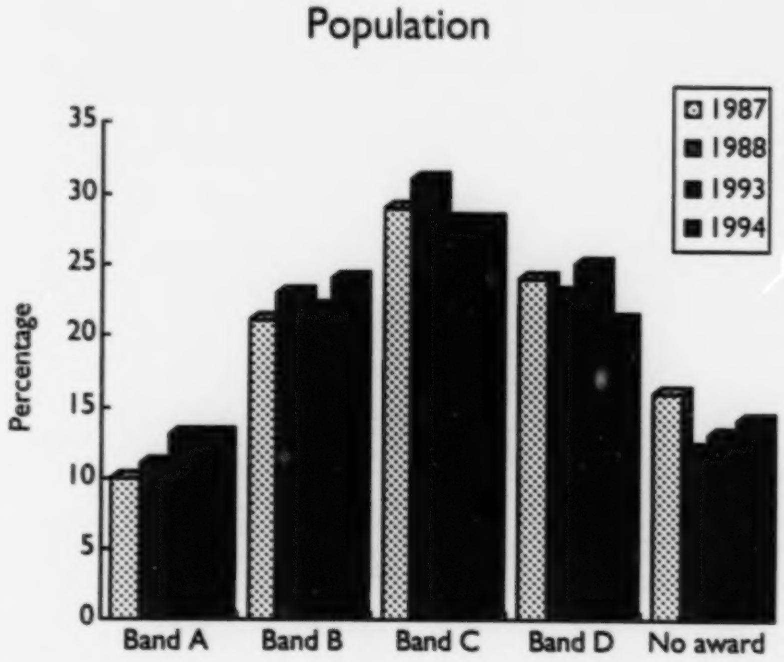
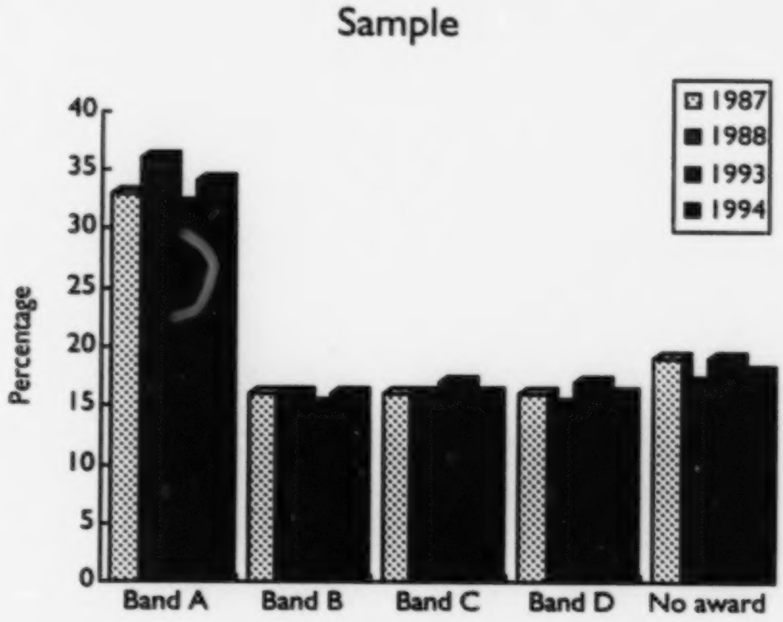


Figure 2.2: Distribution of band awards within the sample available to the project.



Procedure

The researchers, together with HMI and SEB subject specialists, used the 'content and skills' grid, together with markers instructions for the relevant years which the SEB made available to us, to analyse the examination papers. Each question was allocated to a particular cell in the grid. It was then possible to total the scores which candidates had obtained for those questions which related to particular skill areas and content areas. Because each of these scores in each year could have different possible maximum totals (because of differing numbers of questions being asked in each area, and differing marks being awarded to questions) these totals were converted to percentages.

In practice, it proved to be very difficult to conduct analysis on the separate content areas, and only limited comments can be made on them. This was because the division of the data into skill areas and then the further sub-division into content areas often meant that scores for each sub-division were often derived from a very small number of questions. In some years certain areas were not covered at all by questions. For example, we could not identify any knowledge and understanding short answer questions on reproductive biology in 1987, nor any knowledge and understanding short answer questions on adaptation in 1993. The other skill areas had even more gaps. There is therefore, very little that can be said about individual content areas, especially in their 'practical applications' and 'problem solving' aspects.

Findings

Statistical tests were run in order to discover whether there were significant differences between the standards achieved by candidates sitting Higher Grade biology in 1987, 1988, 1993 and 1994. A probability level of 0.01 or smaller was set as the critical level.

It was possible for us to look at performance in each of the skill areas, at each band of award, individually. The mean scores for each skill area, for each band are shown below in Table 2.5, together with the results of tests of statistical difference between the candidates from each year. It is important to note that the statistical tests were not performed on these mean values, but on the raw scores for the individual candidates. They are reported in this table simply for convenience. Mean scores conflate the data and thus leave much out of account. The grouped distributions of these scores in the sample are shown in Appendix B.2.

Table 2.5: Performance in Higher Grade biology skill areas

Skill area	Band	Mean Scores				Statistical
		1987	1988	1993	1994	difference*
Knowledge	A	81.9	82	87.5	87.7	p=0.0005
and	B	62.2	68.8	62.8	69.4	NSD**
Understanding	C	52.8	55	40.8	48	p=0.0003
	D	39.9	46.5	33.2	36.2	p=0.0091
	No award	25.4	31.1	12.4	16.8	p=0.0000
Problem	A	76.9	83.7	74	81.5	NSD
Solving	B	70.4	66.1	56.6	67.6	NSD
	C	61.2	60.6	50	64.9	NSD
	D	46.6	51.4	34	51.6	NSD
	No award	43.2	52	17.4	36	p=0.0000
Practical	A	67.7	82.2	77.2	70.5	NSD
Applications	B	47.9	62	54.7	53.6	NSD
	C	26.9	42.5	33	55.3	p=0.0011
	D	22.8	34.7	28.2	42.9	NSD
	No award	19.1	22.4	14	22.4	NSD

**NB: statistics were calculated from the actual distributions of scores and not the mean values reported in this table.*
***NSD = no significant difference.*

There is a general tendency for candidates who were awarded an overall band A to perform better than those who were awarded band B (and so on through the bands) in the individual skills. In other words there does not appear to be any aberrant behaviour in performance in particular skill areas.

There are a number of statistically significant differences in Table 2.5, some of which appear to reflect the comparatively poor performance of candidates in 1993. Others have different apparent causes. The difference in performance over the years in knowledge and understanding, amongst those candidates who were awarded band A, appears to be caused by higher levels of performance in 1993 and 1994 than in 1987 and 1988. The converse is true for those candidates who were awarded bands C, D and No award: they appear to exhibit poorer performance in knowledge and understanding in 1993 and 1994 than in 1987 and 1988. One hypothetical explanation is that this component of the examination in 1993 and 1994 may have been separating out the different levels of performance more widely than in the earlier years in the sample, perhaps by being more demanding of average candidates while still allowing the more able to demonstrate the full extent of their knowledge and understanding.

However, we cannot ascribe these differences to any general trend across all skill areas. They only appear within the knowledge and understanding component of the examination.

■ The content areas

As we noted earlier, there are few conclusions which can be drawn from the limited data gathered on performance within each of the identified content areas. This is primarily because of the small number of items identified in each area. In particular there were very few problem solving and practical applications questions in any of the content areas. The numbers of knowledge and understanding items in each content area, in each year, are shown in Table 2.6, together with the total number of marks available for correct answers to those items (in brackets).

Table 2.6: Number of knowledge and understanding items identified for each content area in the examination paper for each year

Content Area	1987	1988	1993	1994
Cell biology	5(26)	6(14)	3(21)	5(15)
Genetics and evolution	6(30)	4(12)	3(11)	3(12)
Control of growth and development	1(4)	1(1)	4(13)	2(10)
Regulation in biological systems	0(0)	2(9)	2(5)	2(3)
Adaptation	1(2)	3(9)	0(0)	3(9)

We can see that some content areas in Paper II of the examinations were very lightly assessed in some years. Cell biology and Genetics and evolution appear to be the areas which received most attention, both in terms of the numbers of items asked and of the marks available to be awarded for them (which we can take to be indicative of the complexity and coverage of the individual item). There appears to be some increase in the coverage of Control of growth and development in the later years of our sample. We do not know if this is a trend in the coverage of the examination or merely something peculiar to the particular years included in our sample.

Insofar as we could, we looked for statistical differences in performance in these content areas between years. The numbers of items and marks available within Control of growth and development (in 1987 and 1988), Regulation in biological systems and Adaptation were so low that we must again stress the extreme caution that must be exercised with these findings.

In Cell biology, there was a statistically significant difference between years for candidates who had been awarded bands C and No award (but not other bands). At these bands of award performance was lower in 1993 than in other years. (Again we would remind readers of our caution about any conclusions which are dependent on the small 1993 sample.)

In Genetics and evolution there was a statistically significant difference between years for candidates who were awarded band A (but not other bands). Band A candidates performed better in 1993 and 1994 than in 1987 and 1988.

In Control of growth and development candidates who were awarded band A in 1987 and 1988 exhibited a statistically different (and apparently better) performance than band A candidates in 1993 and 1994, but this is based on only one question in each of the 1987 and 1988 papers and must be regarded as a highly unreliable finding.

Similarly, such apparent statistical differences between years in Regulation in biological systems and in Adaptation as appeared were based on too little data to be regarded as in any way reliable.

The only one of these findings to which we can give any weight is that in Genetics and evolution where the best (band A) candidates have improved their performance in 1993 and 1994 compared to their equivalents in 1987 and 1988.

The essay questions

A more focused study of differences over time was made possible by the existence of the two essay questions which were repeated in different years. The questions themselves have already been quoted (Tables 2.2 and 2.3). Marking schemes for these essay questions were devised by the subject specialists. These marking schemes essentially consisted of checklists of particular items of content which would be expected in answers to the questions (they are presented in Appendix B.1).

The subject specialists then re-marked all the essays according to these marking schedules. A total 'success score' for each candidate was calculated by summing the number of items which they successfully included in their response. It was then possible to test whether there were any differences between performances in the two years in which each essay appeared. The results are shown in Tables 2.7 and 2.8.

It should be noted that the maximum 'success score' which could be awarded by the markers to the candidates in this exercise differed from the maximum number of marks available in the examination because we were using newly created marking schemes. These were specially devised for each question and were designed to be equally applicable to each occurrence of the question in different years.

Table 2.7: Photosynthesis essay, 1988 and 1994 (n = 113)

	Mean scores (Max 21)		Statistical
Band	1988	1994	differences
A	10.9	14.0	p=0.0001
B	8.9	11.8	NSD
C	7.7	7.6	NSD
D	3.1	2.9	NSD
No award	3.2	0.8	NSD

Table 2.8: Industrial melanism essay, 1987 and 1994 (n = 91)

	Mean scores (Max 8)		Statistical
Band	1987	1994	Differences
A	6.1	5.3	NSD
B	5.2	4.0	NSD
C	5.5	4.2	NSD
D	3.8	2.7	NSD
No award	1.8	2.3	NSD

The figures in the tables are mean scores and are not converted to percentages. Statistical differences were calculated on raw scores of candidates (not mean scores of groups) using the Mann-Whitney test for two independent samples.

Again, it is worth noting the obvious correlation between the band of award and the success score in each case. This simply confirms that candidates who are eventually awarded band A tend to do better than those who are awarded band B (and so on through the bands) in these essay questions as elsewhere in the examination. In other words the essays do not reveal any aberrant performance.

The results clearly show that there is no evidence for any change in performance at any band of award between 1987 and 1994 for the industrial melanism essay, and only one apparent change between 1988 and 1994 for the photosynthesis essay. The one difference is that in 1994 those candidates who eventually were awarded a band A exhibited a higher level of performance than did the band A candidates in 1988. This difference is illustrated in the cumulative frequency graphs of Figures 2.3 and 2.4, where we can see that very many more of the 1994 band A candidates are achieving higher scores than their 1988 predecessors.

Figure 2.3: Performance of band C candidates on photosynthesis essay

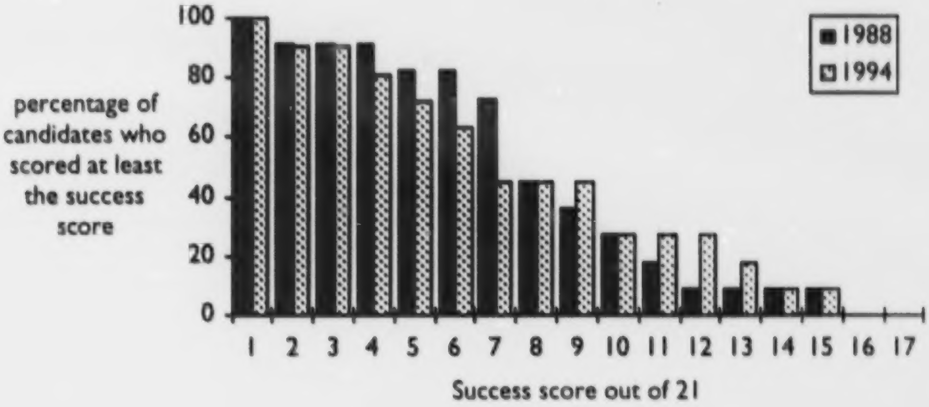
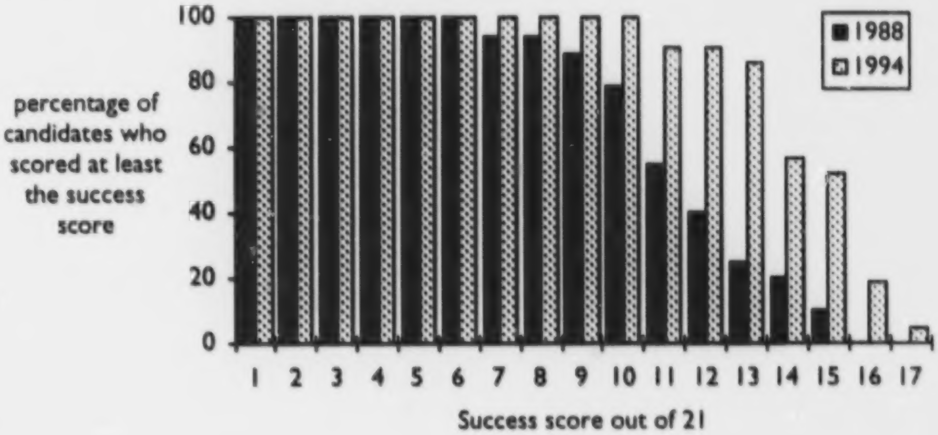


Figure 2.4: Performance of band A candidates on photosynthesis essay



That there is no difference in performance at any band of award below band A suggests that this difference is not caused by any change in 'standards'. Rather it would seem to reflect a genuine improvement in the performance of the very best candidates. This supports the hints of such a change which were also apparent in the performance of 1994 band A candidates in the knowledge and understanding skill area, and in the Genetics and evolution content area.

■ Conclusions

We can conclude the following:

- There was no evidence to support the view that there has been an overall change in standards in Higher Grade biology in the years covered by the study. Some statistically significant differences were observed in performance between years in particular skill areas at particular bands of award. However these did not tend in any one direction or indicate any general shift in standards.
- The more detailed study afforded by the specific examples of the two repeated essays supports the view that there has remained a consistency of standards at almost all bands of award.
- There are some suggestions, though they are not conclusive, that the very best candidates are exhibiting higher levels of performance in the later years of the sample.

If any exercise such as this were to be planned in the future we would recommend that the following points should be considered:

- It would be useful if the Examination Board were to consider their criteria for retention of the sample of scripts. The exact nature of the sample retained would depend on the purpose envisaged for any future exercise, but we can see little merit in the current arrangement which results in great over-representation of band A candidates in the sample. This recommendation applies to all subject areas in the current study.
- Any future exercise such as this would be greatly assisted if it were possible to match candidates performance in Paper I (the item-banked objective test paper) with the archived scripts from Paper II. It would be of greatest use to future research if the actual answer sheets of the sample candidates could be preserved in the archive. Future researchers could then refer back to the appropriate question papers and identify candidates' performances in skill and content areas in Paper I. Some indication of the overall score achieved in Paper I would be the minimum useful information desired.

Appendix - Biology

B.1 Marking schemes for essays

B.2 Graphs illustrating the distribution of scores in skill areas

Appendix B.1

Mark Scheme Essay

Industrial Melanism

Learning Outcomes	Candidate Numbers																				Success score
1. Light and dark forms of the moth																					
2. Pollution/SO_2 kills lichens on the bark/trunk/trees																					
3. Soot darkens the bark																					
4. Moths rest on trunk during the day																					
5. Dark form better camouflaged OR converse																					
6. Fewer of dark form eaten by birds OR converse																					
7. More of dark form survive and breed OR converse																					
8. Increase in number of dark form in industrial areas																					
Individual Success score																					

Mark Scheme Essay

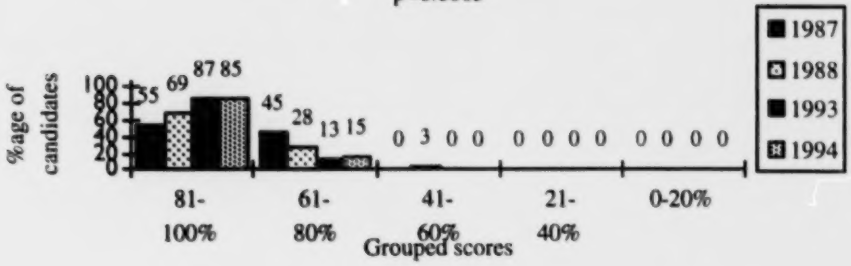
Photosynthesis

Learning Outcomes	Candidate Numbers																				Success score
1. Light stage occurs in the grana																					
2. Pigments/Chlorophyll absorbs light energy																					
3. Several pigments allow absorption over a wide spectrum																					
4. (a) Chlorophyll is the main pigment																					
(b) Chlorophyll absorbs in the red and blue																					
5. Other pigments pass energy to chlorophyll a																					
6. Pigments/Chlorophyll convert light energy to chemical energy																					
7. Energy used to split water/in photolysis																					
8. (a) Hydrogen combines with hydrogen carrier																					
(b) NADP + $\rightarrow$ H_2 $NADPH_2$																					
9. Energy used to synthesis ATP from ADP + P																					
10. Dark reaction occurs in stroma																					
11. Each stage controlled by an enzyme																					
12. RUBP/RUDP is the carbon dioxide acceptor																					
13. A 6C unstable compound is formed																					
14. Two molecules of GP/PGA formed																					
15. The GP/PGA is converted to glucose																					
16. Number of carbons RUBP = 5C GP = 3C Glucose = 6C																					
17. (a) Hydrogen supplied by reduced hydrogen acceptor																					
(b) Hydrogen supplied by $NADPH_2$																					
18. Energy supplied from ATP																					
19. Glucose converted to starch																					
20. RUBP/RUDP reformed from GP/PGA																					
21. Cyclic reaction called Calvin cycle																					
Individual Success score																					

Appendix B.2
Higher Grade biology: distribution of scores in each skill area for each band of award

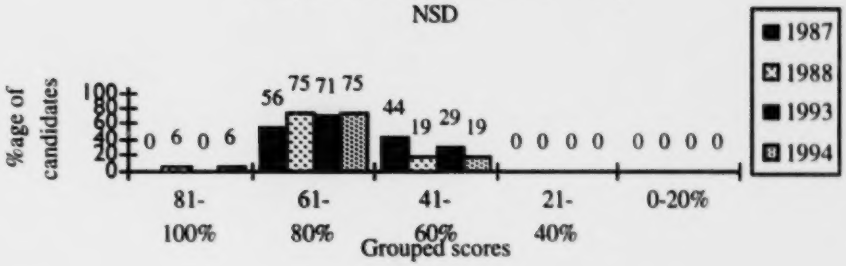
Band A: Knowledge and Understanding

p=0.0005



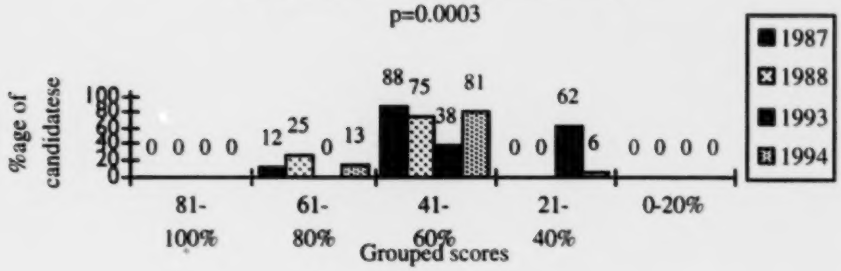
Band B: Knowledge and Understanding

NSD



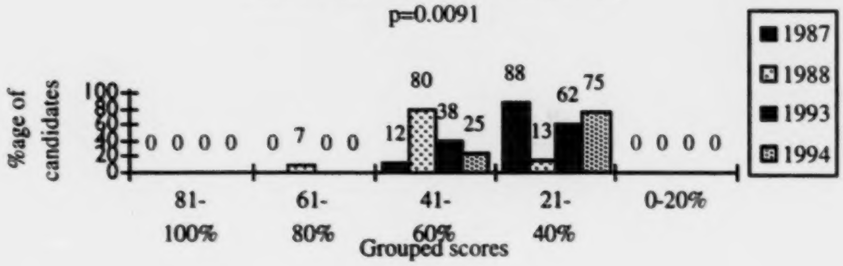
Band C: Knowledge and Understanding

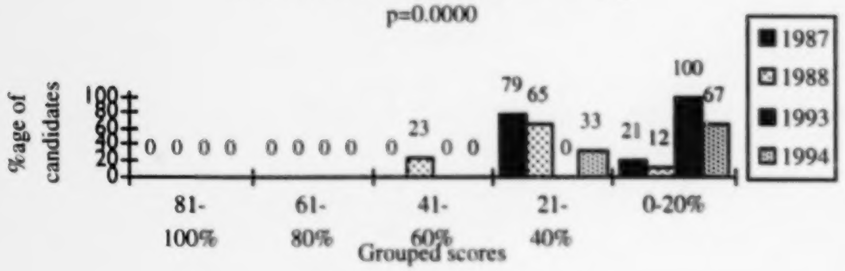
p=0.0003



Band D: Knowledge and Understanding

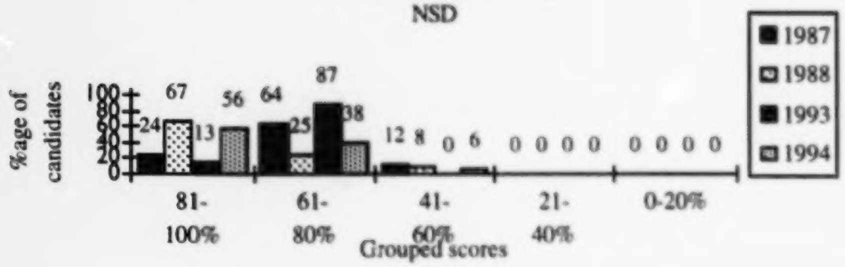
p=0.0091





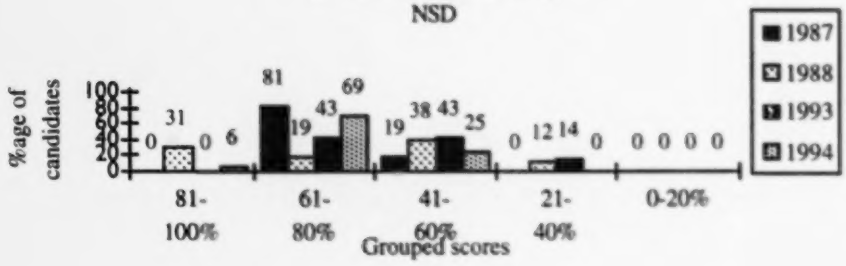
No award: Knowledge and Understanding

p=0.0000



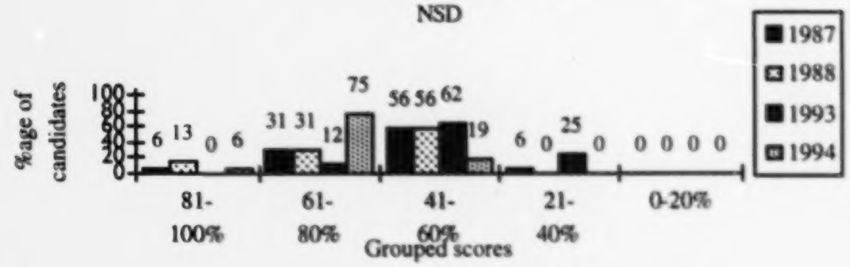
Band A: Problem Solving

NSD



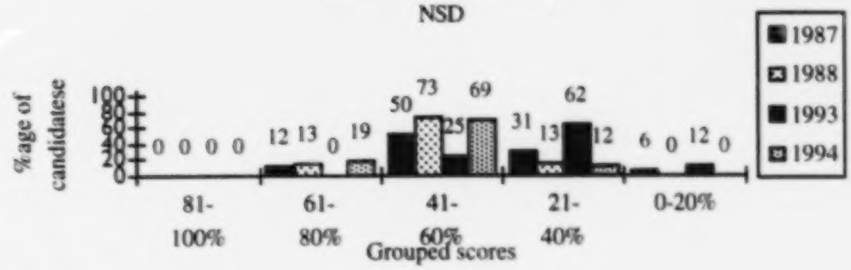
Band B: Problem Solving

NSD

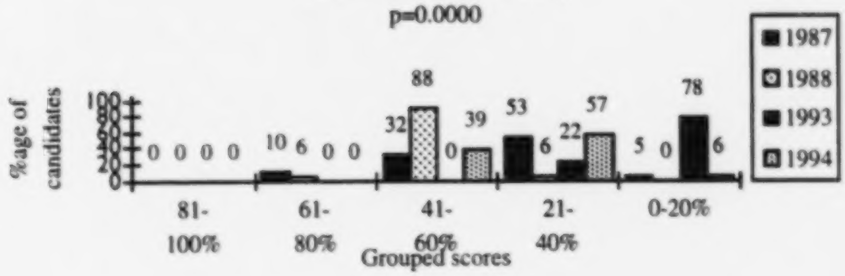


Band C: Problem Solving

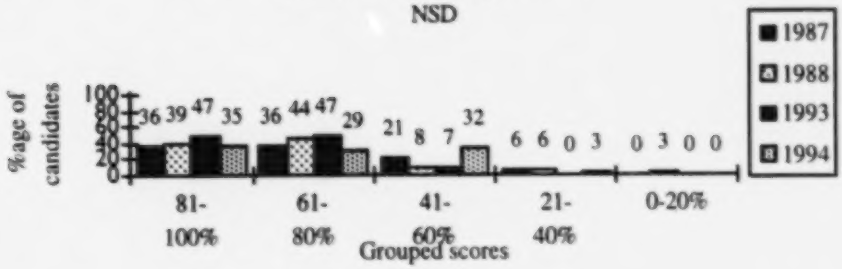
NSD



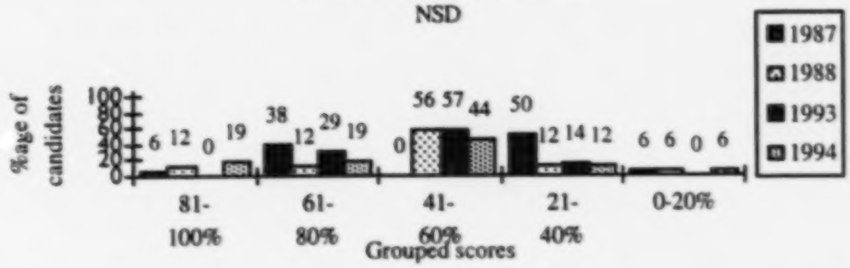
Band D: Problem Solving
NSD



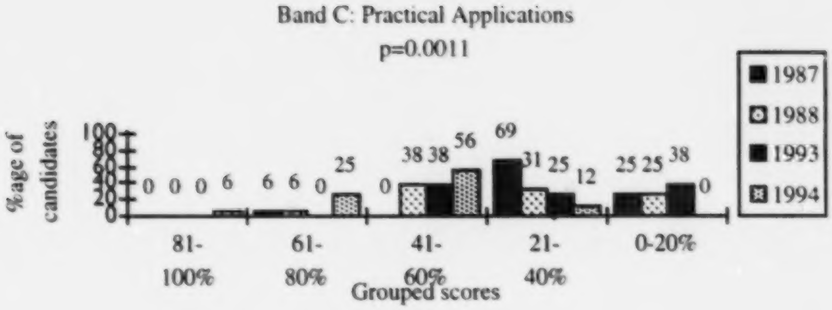
No award:Problem Solving
p=0.0000

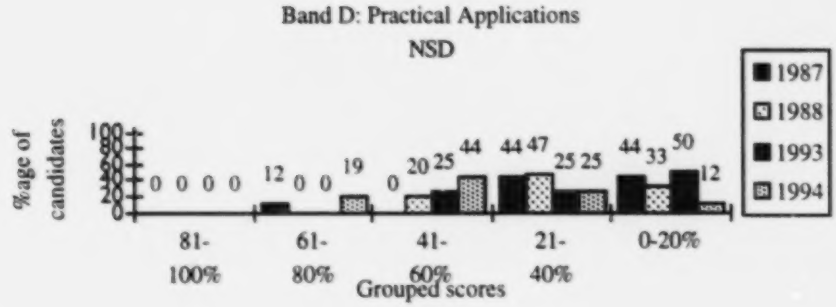


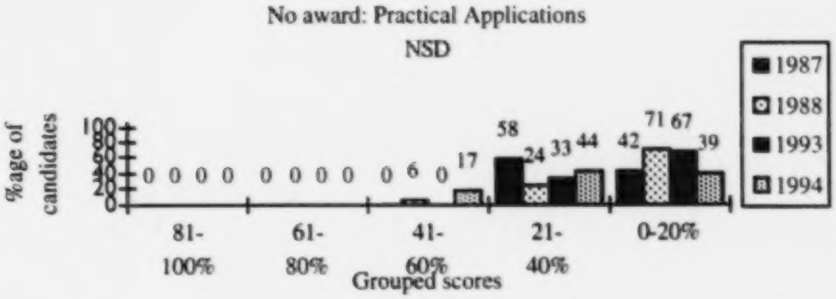
Band A: Practical Applications
NSD



Band B: Practical Applications
NSD


Band C: Practical Applications
p=0.0011


Band D: Practical Applications
NSD


No award: Practical Applications
NSD

3 English

It was clear from an early stage in the project that the core elements of Higher Grade English had remained relatively consistent over the years from which the sample was drawn. This had considerable advantages for the research as there was little need to make allowances for changing coverage between the unrevised and revised Higher Grade examinations. However, because much of the Higher Grade English archive consisted of extended pieces of writing by the candidates, there was a clear need for substantial subject expertise for the re-assessment of candidates' performance. This was available to the research team through the active participation of HMI and SEB personnel, and through the use of a team of experienced SEB markers for the re-assessment of scripts from the archive.

There were, however, some differences between unrevised and revised Higher Grade English syllabuses, and there were also some other factors which meant that the research could not address all aspects of performance in Higher Grade English.

Answering questions on some aspect of the media has been an option within Higher Grade English for many years, but it was felt by our team of subject experts that the ways in which media studies is both taught and assessed had changed considerably between 1987 and 1994. Approaches to this area had become very much more sophisticated than before, and it was therefore thought to be impossible to attempt any meaningful comparison between the early years in our sample and the later years. It was also pointed out to us that, being only an option, very few of the sample scripts would contain media studies answers, and that we would be unlikely to obtain a sufficiently large sample. Media studies answers were therefore excluded from the analysis.

Practical criticism questions had also been options which were little chosen by examination candidates and were unlikely to provide a large sample. These were also excluded from the analysis.

The exact composition of the Higher Grade English examination has changed from unrevised to revised versions. The portfolio approach of the Review of Personal Reading has been introduced, and the 'summary' question in the unrevised examination has disappeared. Nevertheless, there remains considerable commonality.

The main areas which were to be included in the analysis of performance were:

- Literature - evidence for which could be obtained from candidates' reviews of personal reading, critical essays, and some of the more extended responses required in parts of the close reading paper.
- Writing - evidence for which could be obtained from discursive and imaginative pieces of writing by candidates, and also from any reports and/or summaries which they were required to write.

- Interpretation - evidence for which could be obtained from candidates' answers to short response questions in the Interpretation and/or Close reading papers.

Methodology

The sample of scripts from the SEB's archive were re-marked according to strict and uniform criteria to be applied across all sample years where this was possible. This framework ensured a consistent measurement of levels of academic attainment achieved by pupils in 1987, 1988, 1993 and 1994. The use of the same marking scheme for all years (where possible) made it easier to compare performance in different years.

Examination scripts were re-marked by a team of twelve experienced markers chosen in consultation with HMI and the SEB.

Stage 1: Drafting grading schedules

In order to determine what criteria would be used to analyse the various elements of Higher Grade English, a subject group comprising SCRE researchers, members of HMI and subject specialists from the SEB met to discuss the drafting of grading schedules.

To establish a framework for the grading schedules, copies of examination papers used over the 7 year period of our study were obtained, as were SEB-produced extended marking instructions, and copies of the relevant Arrangements documents.

The following principles guided the choice of the criteria to be used in the grading schedules :

- Any analytical framework used should be recognisable to subject experts as a working synthesis of the subject as they teach it.
- It should comprise criteria which are discrete.
- It should achieve a balance between the need to describe detail and the advantages of working with a small number of 'high import' domains.
- It should be sufficiently robust to accommodate changes in the subject over the years covered by the project.

In practice it was found that the existing frameworks included in the SEB Revised Arrangements were applicable to all years of the sample and met all the above criteria. There were one or two minor amendments to the wording of individual statements within these marking grids to take account of slight differences between the unrevised and revised examinations. The marking grids as used are reproduced in Appendix E.1. It can be seen that they are basically criterion-referenced, grade-related statements of performance on sets of subject-related criteria, set at eight levels of performance. In principle they are very similar to the grade-related criteria familiar from Standard Grade examinations.

Such grids were used for all areas of the examination except Interpretation. This is the only area of the Higher Grade English examination which consists of a series of short answer questions and not a piece of extended writing. Because of this the method of marking (and consequently the way that it could be analysed for this research) differed from all other areas. While most areas of the English scripts were given performance marks derived from the grids, and ranging from 1 to 8 (with 1 representing the best performance), in the interpretation paper performance was measured in terms of the number of questions answered correctly (usually converted to a percentage). The questions in the Interpretation paper were also categorised into those addressing a number of sub-skills, from extracting information from a text, through to evaluating an author's use of language (see Appendix E.2 for a full list of these categories).

Stage 2: The re-marking of scripts

As mentioned above, a team of twelve experienced markers was used for the re-marking of all the scripts.

To ensure further checks on the reliability of the re-marking process a training session for markers was held. A set of scripts not included in the formal analysis was used to allow markers to practise marking and ask questions about the process. In addition they were provided with the opportunity to comment on, and if necessary amend the framework of criteria. In this way markers were able to familiarise themselves with the use of the grading schedules.

■ The sample

The sample of scripts obtained from the Scottish Examination Board archive had been based on a random sample controlled by the 'range mark' awarded by the SEB. Since there are more of these 'range marks' than there are final bands awarded in the Higher Grade examination, and since the 'range marks' are not equally mapped onto the final bands awarded (i.e. there are more range marks allocated to band A than to other bands), the structure of the sample does not reflect the distribution of bands in the population.

The SEB retain a sample of 100 scripts each year. These were made available to us for the years 1987, 1988 (the unrevised, traditional examination), 1993 and 1994 (revised Higher). However, some scripts were used in the training of the markers, and others were subsequently used in a markers' trial (see below). This reduced the number of scripts available for analysis. The final sample is shown in Table 3.1.

Table 3.1: Higher Grade English: Sample of scripts

	A	B	C	D	No Award	Total
1987	33	17	14	14	17	95
1988	34	16	16	14	16	96
1993	33	15	15	15	18	96
1994	32	15	16	15	18	96
Total	132	63	61	58	69	383

Trial marking

Twelve individual markers working independently of each other were asked to re-mark identical sets of papers. The scripts of three candidates were presented to each marker. Each paper was to be assessed using the appropriate marking grid, which resulted in a record of performance on each criterion for each piece of work from each candidate. A numerical grading system of 1 to 8 (with '1' representing the best performance) was used. The marking was therefore criterion-referenced using a system analogous to grade-related criteria.

The statistic used to check the inter-rater reliability of the markers was the Kendall coefficient of concordance. This gives a measure of the extent to which the twelve markers were making the same judgements about these scripts when measured against these criteria. It produces a statistic (W) which varies in value between 0 and 1. A value of 0 represents a complete lack of agreement between markers, while a value of 1 represents complete agreement. The test produced a value of W=0.8185, with a level of statistical significance which implied that such a value could have had less one chance in ten thousand of having occurred by chance. This represents a high level of agreement between the markers, with an extremely low probability that such agreement could have been arrived at by chance.

As a check we also looked at the individual correlations between markers, using Spearman's correlation coefficients. Again these showed high levels of agreement, with correlation coefficients in the region of 0.8. The probabilities of these levels of agreement arising by chance were less than one in a thousand.

The conclusion from both these tests is that there was a good level of inter-rater reliability, and we can have confidence in the consistency of the assessments being made by the group of markers.

The component parts

The study looked at the following components within the Higher Grade English papers:

- Discursive writing (including personal reflection)
- Imaginative writing
- Reports and summaries
- Critical essays and Review of Personal Reading (RPR) pertaining to - Drama
- Critical essays and Review of Personal Reading (RPR) pertaining to - Prose
- Critical essays and Review of Personal Reading (RPR) pertaining to - Poetry
- Critical essays and Review of Personal Reading (RPR) pertaining to - Folio (in 1993 and 1994 only)
- Interpretation

It was decided to exclude the following components from the study:

- Media studies - as a relatively new addition to the Higher Grade English syllabus it was constantly undergoing change. Tended not to be done by large numbers of candidates
- Practical criticism - partly covered by generic criteria. Tended not to be done by large numbers of candidates.

The components to be included in the study, bar 'interpretation', were broken down by criteria and marked as described above. These components (Discursive writing; Imaginative writing; Reports and summaries; Critical essays and Review of Personal Reading (RPR) pertaining to drama, prose and poetry) were treated separately. The areas covered by the criteria are listed below. Each criterion had eight levels of performance specified. These are set out in detail in Appendix E.1.

Criteria to be used for marking - Discursive writing (including personal reflection)

Language
Line of Thought
Personal Stance towards Topic
Other Points of View

Criteria to be used for marking - Imaginative writing

Language
Specific Literary Form
Tone/Mood
Thematic Content

Criteria to be used for marking - Reports and summaries

Understanding
Selection
Order and Structure
Formal Tone and Register
Language Competence

Criteria to be used for marking - Critical essays and Review of Personal Reading (RPR) pertaining to Drama

Personal Stance
Relevance/Line of Thought
Knowledge of Text
Textual Analysis Style/Structure
Linguistic/Literacy Awareness
Language Competence

Criteria to be used for marking - Critical essays and Review of Personal Reading (RPR) pertaining to Prose

Personal Stance
Relevance/Line of Thought
Knowledge of Text
Textual Analysis Style/Structure
Linguistic/Literacy Awareness
Language Competence

Criteria to be used for marking - Critical essays and Review of Personal Reading (RPR) pertaining to Poetry

Personal Stance
Relevance/Line of Thought
Knowledge of Text
Textual Analysis Style/Structure
Linguistic/Literacy Awareness
Language Competence

During the training session for the markers they expressed some dissatisfaction with the clarity of differentiation between levels of performance (as defined in Appendix E.1). Close examination of the wording led them to conclude that some definitions were vague and unhelpful. Extensive discussion of their possible meanings, and some slight amendments to wording, did help. As we have reported above, the markers were able to use these criteria, in the practical marking trial, with a satisfactory degree of consistency.

In the case of Interpretation where candidates were asked to respond to a series of short answer questions, it was decided that it would be best to analyse the individual questions and relate each to the following criteria:

Criteria relating to understanding:

- Selecting, explaining and, where appropriate, summarising the main concerns of the text(s) (or major parts of text) and showing their relationships
- Selecting and explaining the meanings of key words and phrases and their relationships
- Drawing inferences from a key statement or statements
- Substantiating statements, explanations and inferences by reference to the text(s).

Criteria relating to appreciation and response:

- Appreciation and response by - making relevant comments on aspects of the point of view of the author(s)
- Appreciation and response by - substantiating comments from evidence in the text(s) and/or from their own experience and knowledge
- Appreciation and response by - analysing and commenting on the use of conventions of the genre, including the structure, style and language of the text(s), applying whatever terminology is appropriate.

We therefore have 31 criteria, each defined at eight levels of performance, plus the seven criteria relating to Interpretation, giving us a total of 38. There is an argument that the criteria used to mark prose, poetry and drama critical essays are identical (although applied to different genres), which would give us a total of only 26 criteria, but the results for each genre are reported separately in Appendix E.3.

■ The findings

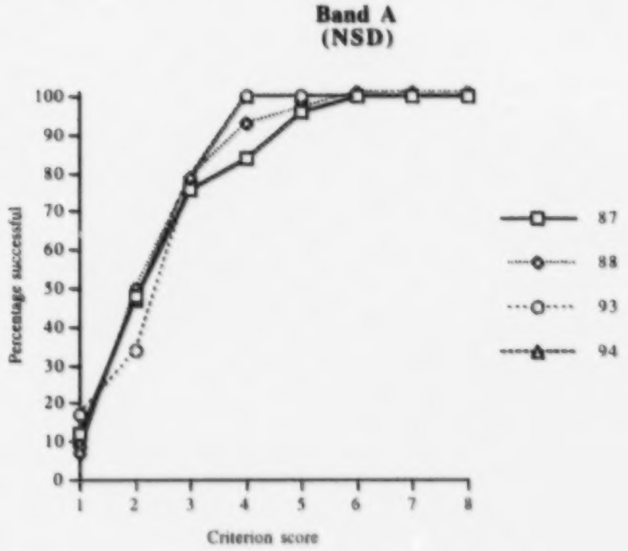
Statistical tests were run to discover whether there were significant differences between the standards achieved by pupils sitting the Higher Grade English papers in 1987, 1988, 1993 and 1994. A probability of 0.01 or smaller was chosen as the level of statistical significance.

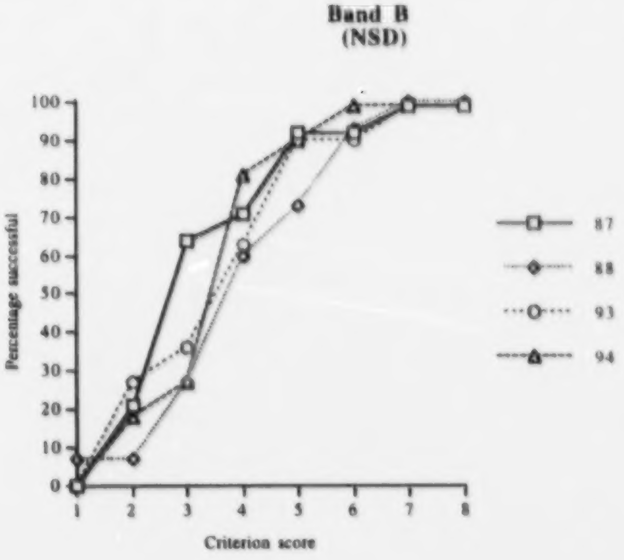
Individual criteria

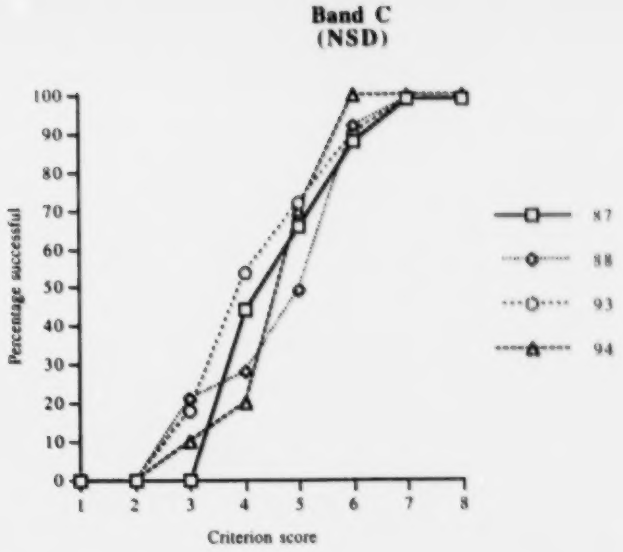
For each band of award (five in all) we examined each criterion, and tested to see if there was any statistically significant difference between performance in each year of the sample. This gives us 5 x 38 = 190 comparisons in all. This is far too many to present here. A complete list, together with the result of the statistical tests (Kruskal-Wallis) is given in Appendix E.3. In all, fourteen statistically significant differences between years were identified. However, only two of these differences occurred outside the Interpretation paper.

We can illustrate a typical set of findings graphically (see Figure 3.1). This shows the cumulative frequency graphs of performance by candidates who were awarded each band of pass, for all four years of the sample. In this case the graphs are of performance in the Language skills criterion of the Discursive writing element of Higher Grade English, but similar patterns were evident almost everywhere. We can immediately see that there is very little separation between the samples from the different years.

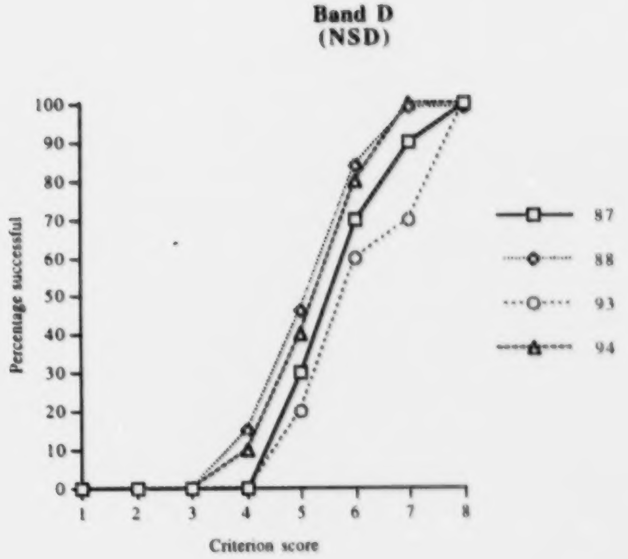
Figure 3.1: Discursive writing (including personal reflection)
Examining language skills







Band C
(NSD)



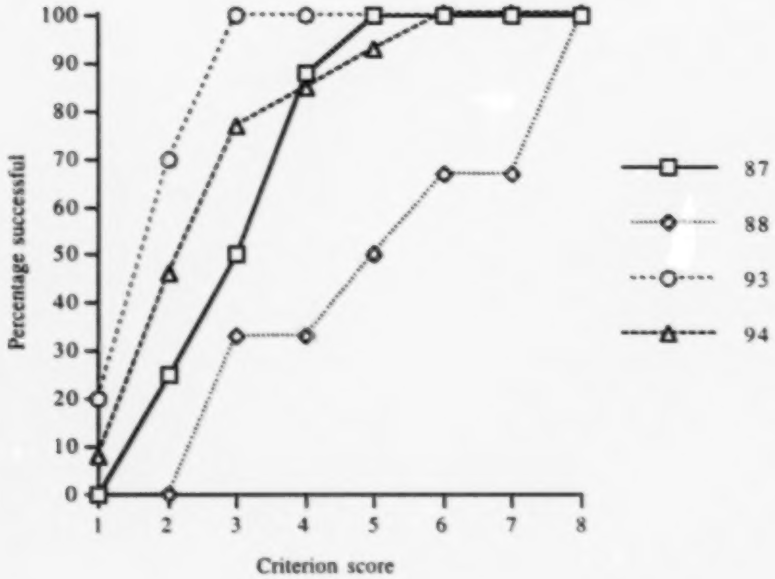
Band D
(NSD)

Overall, the findings showed that there were almost no statistical differences between candidates' achievement levels in years 1987, 1988, 1993, 1994. As mentioned, there were only two exceptions to this outside of the area of Interpretation.

Both instances occurred in the Imaginative writing question: significant differences emerged between band A students in relation to Language (Figure 3.2) and Tone/Mood (Figure 3.3).

These shall be treated in turn.

Figure 3.2: Imaginative writing - Band A
Language criterion (SD)

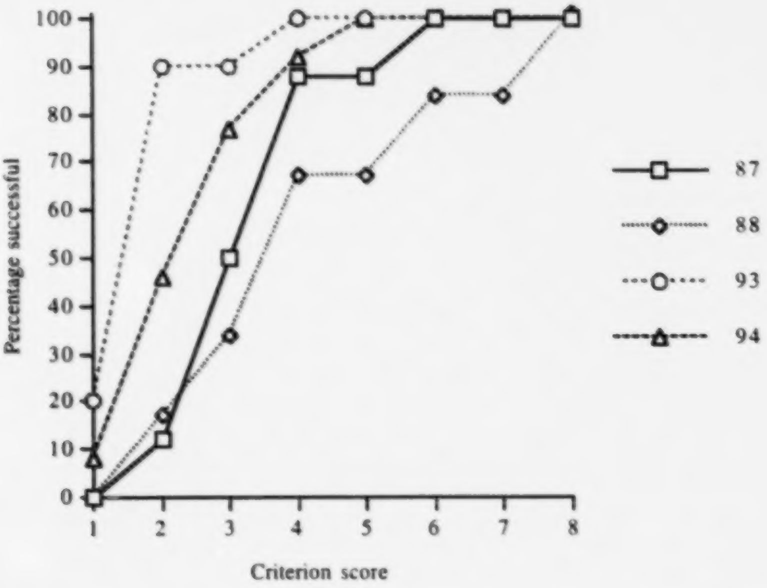


Band A candidates from 1993 performed significantly better than 'A' students from the other three years. In 1993 100% of those candidates who achieved an overall band A award performed at performance level three. That same level of performance was achieved by only (approximately) 50% of band A candidates in 1987, 35% in 1988, and 75% in 1994. However, Imaginative writing is not an option chosen by most candidates and this finding is based on a relatively small sample of 37 candidates in all, split between years as follows:

Year	1987	1988	1993	1994
Number	8	6	10	13

This small sample must cast severe doubt on this finding.

Figure 3.3: Imaginative writing - Band A
Tone/Mood criterion (SD)



For the Tone/Mood criterion it was again band A candidates from 1993 and 1994 who outperformed those from 1987 and 1988. The same proviso as before applies to this finding. With such small numbers in the sample, it is very difficult to give any weight to this finding.

Within Interpretation there were twelve statistically significant differences on performance in individual criteria. Candidates' scores are summarised in Table 3.2 which shows mean scores for each criterion at each band of award for each year, together with the results of the statistical tests (note that the statistical tests were carried out on the raw scores of the candidates and not on the mean scores reported in this table. The two are reported together here for the sake of brevity).

Table 3.2: Higher Grade English: Interpretation

		Mean Scores				Statistical
Criterion	Band	1987	1988	1993	1994	Difference
Interpret 1	A	68.2	76.7	82.8	65.7	NSD
	B	63.2	65.3	66.3	54.5	NSD
	C	53.6	59.3	60.4	36.7	NSD
	D	44.6	43.9	57.5	27.5	p=0.0078
	No award	35.3	34.6	40.9	15.3	p=0.0038
Interpret 2	A	73.0	73.1	80.3	73.1	NSD
	B	62.9	63.7	70.6	54.8	NSD
	C	50.4	56.6	64.1	50.0	NSD
	D	38.1	41.2	47.4	43.2	NSD
	No award	31.2	34.1	39.5	29.7	NSD
Interpret 3	A	82.8	75.9	75.4	62.5	NSD
	B	66.5	62.5	55.0	39.3	NSD
	C	58.3	51.9	35.0	48.1	NSD
	D	50.7	47.8	23.3	28.3	NSD
	No award	43.4	35.7	20.1	23.4	NSD
Interpret 4	A		64.4	78.6	76.4	NSD
	B		53.1	68.4	61.7	NSD
	C		26.6	64.7	52.1	p=0.0015
	D		30.4	51.3	47.7	NSD
	No award		14.1	39.2	33.1	p=0.0024
Interpret 5	A	60.5	41.1	61.1		NSD
	B	48.5	18.8	45.8		p=0.0050
	C	39.3	6.2	33.3		p=0.0002
	D	18.8	3.8	28.6		p=0.0007
	No award	20.0	0	22.9		p=0.0000
Interpret 6	A	48.5	60.7	65.9	55.6	NSD
	B	42.2	40.8	50.4	39.3	NSD
	C	28.6	23.4	30.8	30.4	NSD
	D	15.3	24.0	26.1	16.7	NSD
	No award	12.2	9.4	21.6	3.1	p=0.0002
Interpret 7	A	67.3	63.4	67.9	65.3	NSD
	B	60.0	40.4	52.8	48.9	p=0.0024
	C	49.2	27.3	39.2	38.6	p=0.0030
	D	31.8	25.8	31.5	29.2	NSD
	No award	29.8	12.2	23.5	14.2	p=0.0000

All figures are mean scores of candidates in each year, at each band of award, converted to percentages. Higher numbers represent better performance. This is the opposite to the marking system for other aspects of Higher English, where lower numbers represent better performance.

Key to Criteria

Criteria relating to understanding:

Interpret 1: Selecting, explaining and, where appropriate, summarising the main concerns of the text(s) (or major parts of text) and showing their relationships.
Interpret 2: Selecting and explaining the meanings of key words and phrases and their relationships.
Interpret 3: Drawing inferences from a key statement or statements.
Interpret 4: Substantiating statements, explanations and inferences by reference to the text(s).

Criteria relating to appreciation and response:

Interpret 5:	Appreciation and response by making relevant comments on aspects of the point of view of the author(s).
Interpret 6:	Appreciation and response by substantiating comments from evidence in the text(s) and/ or from their own experience and knowledge.
Interpret 7:	Appreciation and response by analysing and commenting on the use of conventions of the genre, including the structure, style and language of the text(s), applying whatever terminology is appropriate.

The performance scores for Interpretation differ from the other performance scores obtained by the research in some ways. Instead of being expressed in terms of eight levels of performance (with '1' representing the best performance), they are expressed in terms of the number of questions correctly answered, converted to a percentage to allow for comparisons between samples where there were different numbers of questions asked for each criterion. The higher numbers therefore represent the better performance. It is also the case that not all criteria were covered by questions in all years. Interpret 4 was not included in the 1987 paper, and Interpret 5 was not included in 1994.

Two possible explanations for these differences can be suggested, but cannot be conclusively shown to be their cause. One is that there are differing numbers of questions addressing each of the criteria in each year, and it may not always be the case that all criteria are adequately covered (some do not figure at all in some years). The numbers of items which were identified as addressing each of the Interpretation criteria in each paper are shown in Table 3.3. Note that it was possible for any one item to be classified as addressing more than one of these criteria. The actual classification of individual items is given in Appendix E.2.

Table 3.3: Total number of items addressing each criterion in the Interpretation papers.

Criterion	1987	1988	1993	1994
Interpret 1	1	5	4	2
Interpret 2	13	7	8	5
Interpret 3	5	3	1	1
Interpret 4	0	1	3	2
Interpret 5	2	1	1	0
Interpret 6	1	3	2	1
Interpret 7	10	6	5	8

We can see that the numbers of items which address the criteria Interpret 3, 4, 5 and 6 are particularly low, and that the first criterion varies in the number of items identified with it from year to year. Given such small numbers of items the scope for apparent fluctuations in performance from year to year is increased. The force of this argument is increased when we consider that analysis of the candidates' overall performances in the Interpretation papers (i.e. looking at their total score for Interpretation and not the performance on individual criteria) showed that there were no statistically significant differences in performance between years at any band of award.

The second possible explanation is that the level of difficulty of individual questions addressing specific criteria differs from year to year (overall in any one year this need make little difference as easy questions in one area will be balanced by more difficult questions in another, but it could affect results in an exercise like this where the focus is on the individual criteria in isolation). There is some support for this explanation in item difficulty statistics which the SEB supplied us for the years covered by these samples. However, the exact statistics calculated by the SEB have changed over the last few years and proper conclusions cannot be drawn from those available. The small number of items addressing some of the criteria would exacerbate this problem of differences in item difficulty.

We must therefore be very wary of drawing any conclusions from these apparent differences in the candidates' performances in the individual criteria in the Interpretation papers because:

- there may not be a sufficient number of items addressing each criterion, and
- the difficulty of items addressing individual criteria may vary from year to year, therefore
- the level of analysis may be too fine (individual criteria instead of overall score on the paper).

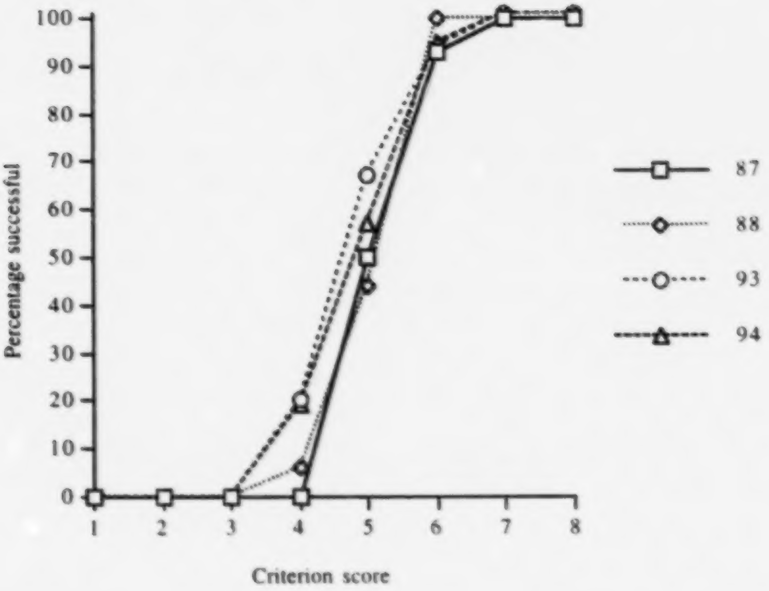
If it is accepted that the interpretation of written text is itself a single, discrete skill (which, although analysable into the constituent parts represented by the individual criteria, is greater than the sum of these parts) then we should draw reassurance from the lack of any statistically significant difference in the overall scores of candidates between years and conclude that levels of performance have not changed. If, on the other hand, it is maintained that the individual criteria represent different skills, which are distinct from one another, and should not be amalgamated in this way, then there is little that we can conclude because of the insufficiency of the data on individual criteria.

Language skills

'Language skills' are a very important part of Higher Grade English and appear as a specific criterion in various places throughout the examination. These include: the discursive essay, imaginative writing, reports and summaries, literature (drama), literature (prose), literature (poetry), RPR and folio. An amalgamated score was created from all these sources for each candidate and statistical tests were run to compare performance on this criterion in different years.

No statistically significant differences between years were found. Typical performance across years is illustrated graphically below for candidates who were awarded band C.

Figure 3.4: Language skills - Band C (NSD)



Conclusions

There are two main conclusions to be drawn from these findings:

- There is no evidence for any overall variation in standards in the years covered by our sample in Higher Grade English. The performance represented by each band of award appears to remain consistent.
- The very few differences in performance on individual criteria which were found were not indicative of any general pattern and were likely to be explained by features of the sample or the particular papers set in each year.

Appendix - English

E.1 Marking grids used in the research

E.2 Criteria for categorising interpretation questions

E.3 List of criteria compared across years together with results of statistical tests

SCRE Project: Higher English

Appendix E.1

Criteria to be used for re-marking discursive writing (including personal reflection)

CRITERIA	CATEGORY 1	CATEGORY II	CATEGORY III	CATEGORY IV	CATEGORY V	CATEGORY VI	CATEGORY VII	CATEGORY VIII
LANGUAGE 1	Fluent, original & stylish, exploiting language resourcefully and subtly to achieve desired effect	Used confidently, accurately and skilfully to achieve desired effect	Used confidently, accurately and appropriately to achieve deliberate effect	Used in a sufficiently varied and accurate way to communicate effectively	Satisfies the standards of formal competence though lacks variety. Communicates clearly enough	Just satisfies the standards of formal writing competence, with occasional lapses and inaccuracies	Just fails to satisfy the standards of formal writing competence	Lapses and inaccuracies are too frequent to satisfy the standards of formal competence
LINE OF THOUGHT 2	Convincing and sustained, revealing a clarity of insight	Sound and sustained and reveals insight into the topic	Clear and sustained though not fully developed	Clear, even if sustained in a pedestrian manner	Can be recognised despite some gaps, inconsistencies or irrelevance in the argument/evidence presented	Can be recognised though ideas are immature, or underdeveloped, with gaps, inconsistencies or irrelevance	Ideas and evidence are thin	Hardly discernible
PERSONAL STANCE TOWARDS TOPIC 3	Original point of view skilfully, imaginatively and convincingly sustained	Reveals commitment and originality	Shows some commitment and some originality	Some signs of a genuine personal stance	Discernible, but shows few signs of genuine commitment: not confidently sustained	Not particularly original, nor confidently sustained, but still discernible	Predictable and/or inconsistent	Hardly discernible
OTHER POINTS OF VIEW 4	Skilfully exploited in the writer's line of argument	Given due weight and used skilfully in the line of argument	Incorporated in the argument	Recognised, though not given due attention in the line of argument	Awareness is shown, though handled without confidence	Awareness is shown, though handled artlessly	Awareness exists but ideas hardly integrated in the argument	Very little is shown of other points of view

SCRE Project: Higher English
Criteria to be used for re-marking reports and summaries

CRITERIA	CATEGORY 1	CATEGORY II	CATEGORY III	CATEGORY IV	CATEGORY V	CATEGORY VI	CATEGORY VII	CATEGORY VIII
UNDERSTANDING 1	Intelligent understanding of task and material	→	Lucid understanding of task and material	Understanding of task and material	Basic understanding of task and material	Basic understanding of task and material	Flawed understanding of task and material	Little understanding of task and material
SELECTION 2	All relevant material selected and skilfully synthesised	All relevant material selected and fully synthesised	Material selected is relevant but not fully synthesised	Material selected is relevant though there may be little synthesis	Most of the relevant material there, though there may be minor omissions	Most of the relevant material there, despite minor omissions or misunderstandings	Some of the relevant material selected but some omissions and misunderstandings	Selection based on misunderstanding or is hardly attempted
ORDER AND STRUCTURE 3	Material and ideas handled confidently and skilfully ordered in a clear and logical structure	Material and ideas handled confidently in a logical structure	Material and ideas ordered effectively to provide a logical sequence	Material and ideas structured to give a generally logical sequence	Thematically connected items are grouped in an appropriate sequence with some indication of relative importance	Thematically connected items are grouped to provide a loose but recognisable structure, but some items out of sequence	Thematically connected items loosely grouped, and some indication of sequence	Little sign of structure or appropriate sequence
FORMAL TONE AND REGISTER 4	Stylish and confidently maintained	Confidently maintained	Maintained with competence	Maintained	Maintained with few lapses	Unevenly maintained	Barely maintained	An inappropriate tone or register is used
LANGUAGE COMPETENCE 5	Reads well with variety of sentence structure, good linkage and individual style	→	Reads well enough, sound sentence structure, competent linkage	Reads well enough, sentencing basic but mainly correct, sensible linkage	Basic, occasionally flawed linkage, a monotonous sentence structure	Basic, occasionally flawed linkage and sentence structure	Poor linkage, weak sentencing	Little or no linkage, severe communication problems

SCRE Project: Higher English
Criteria to be used for re-marking imaginative writing

CRITERIA	CATEGORY I	CATEGORY II	CATEGORY III	CATEGORY IV	CATEGORY V	CATEGORY VI	CATEGORY VII	CATEGORY VIII
LANGUAGE 1	Fluent, original & stylish, exploiting language resourcefully and subtly to achieve desired effect	Used confidently, accurately and skilfully to achieve desired effect	Used confidently, accurately and appropriately to achieve deliberate effect	Used in a sufficiently varied and accurate way to communicate effectively	Satisfies the standards of formal competence though lacks variety. Communicates clearly enough	Just satisfies the standards of formal writing competence, with occasional lapses and inaccuracies	Just fails to satisfy the standards of formal writing competence	Lapses and inaccuracies are too frequent to satisfy the standards of formal competence
SPECIFIC LITERARY FORM 2	Accomplished use of the chosen literary form exploiting the genre creatively	Exploitation of the chosen form is skilful, and the piece is well-structured	Handling of the chosen form is effective, with a clear structure	Some understanding in handling the form is displayed: the structure is adequate if not entirely satisfactory	Some attempt made to employ the conventions of the chosen form. Structure is flawed, but exists.	Awareness shown of some features of the chosen form but their use is structurally flawed or minimal	Some of the relevant features selected but not exploited to any purpose	Does not show awareness of essential features of the form and structure is weak or non-existent
TONE/MOOD 3	Tone/mood controlled, effective skilfully developed and sustained	Tone/mood created effectively and consciously and is sustained	A deliberate tone/mood is consciously created and sustained	A mood/tone is realised though perhaps not sustained and/or developed	A mood/tone is achieved at least at points, even if not very effectively	A discernible attempt to create a mood/tone at some point(s) in the piece	Little attempt to establish mood or tone	No attempt to establish mood or tone
THEMATIC CONTENT 4	Original and well-developed; offering real insights in a challenging or imaginative way	Strong and developed, reaching a satisfying conclusion or offering clear insight	Competently incorporated, communicating personal involvement (where appropriate)	Recognisable and communicates a personal stance/style	Some ideas or comments indicate a personal stance/style but are too disconnected to form a theme	Some ideas or comments indicate a personal stance/style, but writing is mainly literal or 'anecdotal'	General thinness of content and ideas	Little evidence of any theme

SCRE Project: Higher English
Criteria to be used for re-marking critical essays

CRITERIA	CATEGORY I	CATEGORY II	CATEGORY III	CATEGORY IV	CATEGORY V	CATEGORY VI	CATEGORY VII	CATEGORY VIII
PERSONAL STANCE 1	Genuine, personal contribution to the discussion of the topic	Clear personal involvement, confident and committed stance	Consistent personal interest evident and stance shows some commitment	Consistent personal interest is shown but stance may be derivative or mundane	Personal interest/ and/or engagement with text shown, even though unoriginal or intermittent	Originally discernible, though it may be derivative or unoriginal	Where discernible it may be unconvincing	Very little evidence of personal stance discernible
RELEVANCE/ LINE OF THOUGHT 2	Cogent and relevant line developed in a thoughtful way to a genuine and insightful conclusion	Clear and relevant line developed in a thoughtful way to a satisfactory conclusion	Line which has relevance is developed and sustained, though conclusion reached may be modest or limited	A line is offered but not fully developed and/or sustained, and may not reach a satisfactory conclusion	Line of thought is stated but is weak and/or partially dependent on narrative	Line of thought is minimal and/or dependent on narrative. Argument gives some structure but not explicitly stated	Line of thought is thin or almost entirely dependent on narrative. Little structure and no explicit argument	Very little, or only implicit, line of thought can be found
KNOWLEDGE OF TEXT 3	Thorough and perceptive; permeates the answer	Thorough and perceptive	Accurate though literal, supported by apt use of textual evidence	Literal interpretation, supported by sound use of textual evidence	Very basic commentary, supported by adequate use of summary or description of text	An elementary knowledge of at least one central concern of the text is outlined	Sketchy, or confused knowledge or understanding of the text is evident	Knowledge or understanding is hardly shown, or the writing is highly derivative
TEXTUAL ANALYSIS STYLE/ STRUCTURE 4	Knowledge and understanding of effect of style/structure central to argument	Ability to discuss effect of style/structure of text is shown	Awareness of the effect of style/structure of text is shown	Some awareness of style/structure of text is shown	Awareness of structure/style is elementary and lacking in precision	There is some attempt to comment on style/structure	There is a weak attempt to comment on style/structure	Very little evidence of concern with style/structure
LINGUISTIC/ LITERACY AWARENESS 5	Where appropriate, handling of concepts and their effects is subtle and original	Where appropriate, ability to handle concepts and their effects is clearly revealed	Where appropriate, a knowledge of concepts and their effects is revealed	A basic knowledge of concepts (literary/mass media) as appropriate, is revealed	Awareness of concepts, as appropriate is elementary or lacking in precision	Where appropriate, some attempt at relevant reference reveals very basic/implied awareness	Some confused attempts at relevant reference may be evident	Very little awareness of concepts is shown
LANGUAGE COMPETENCE 6	Fluent, original and stylish; exploiting language resourcefully	Used confidently, accurately and skilfully	Used confidently, accurately and appropriately	Used in a sufficiently varied and accurate way to communicate effectively	Satisfies the standards of formal competence though lacks variety. Communicates clearly enough	Just satisfies the standards of formal writing competence, with occasional lapses and inaccuracies	Just fails to satisfy the standards of formal writing competence	Lapses and inaccuracies are too frequent to satisfy the standards of competence

Appendix E.2

Criteria used to classify interpretation questions in Higher Grade English

Close reading of unseen prose, poetry or drama (interpretations)

In Close reading of unseen prose, poetry or drama candidates demonstrate by written answers to specific questioning,

understanding, by some combination of the following:

i. selecting, explaining and, where appropriate, summarising the main concerns of the text(s) (or major parts of text) and showing their relationships;
ii. selecting and explaining the meanings of key words and phrases and their relationships;
iii. drawing inferences from a key statement or statements;
iv. substantiating statements, explanations and inferences by reference to the text(s);

and

appreciation and response, by some combination of the following:

v. making relevant comments on aspects of the point of view of the author(s);
vi. substantiating comments from evidence in the text(s) and/or from their own experience and knowledge;
vii. analysing and commenting on the use of conventions of the genre, including the structure, style and language of the text(s), applying whatever terminology is appropriate.

[The above criteria are drawn from the SEB Revised Arrangements for Higher English, page 22, paragraph 3.4, but slightly modified to suit the purposes of the research].

Classification of questions in interpretation papers

1987		1988		1993		1994	
Paper II		**Paper II**		**Paper I (Rev)**		**Paper I (Rev)**	
Question	**Category**	**Question**	**Category**	**Question**	**Category**	**Question**	**Category**
Question A		Question A		Passage 1		Passage 1	
Section 1		Section 1					
1 a	ii	1 a	iii	a	vii	a	i
1 b	ii	1 b	ii	b i	ii	b	ii
Section 2		1 c	ii/ iii	b ii	i/ ii	c	vii
2 a	i/ vii	Section 2		c	iii/ vi/ vii	d	vii
2 b	ii/ iii	2 a	i	d i	vii	e	ii
2 c	iii	2 b	i	d ii	vii	f	ii
2 d i	ii	2 c	iv	e	ii	g	ii/ vii
2 d ii	ii/ vii	2 di	ii	f	i/ iv	h	iv
2 e	ii/ vii	2 d ii	ii	Passage 2		i	vii
2 f	iii	2 e	i/ ii	g	ii	j	i/ vii
2 g	ii	2 f	ii	h	i/ ii	Passage 2	
2 h i	ii/ iii	2 g	i/ ii	ii	ii	k	ii/ vii
2 h ii	ii/ iii	Section 3	n/a	i ii	i	l	vii
Section 3	n/a	Section 4		j	ii/ iv	m	vii
Section 4		4a	vii	Both		n	iii/ vii
4 a	vii	4 b	vi/ vii	k	ii/ iv	Both	
4 b	vii	4 c	vii	l	v/ vi/ vii	o	vi/ vii
4c	v/ vii	4 d	v/ vii			p	iv
4 d	ii/ vii	Question B					
4 e	v/ vi/ vii	a	i				
Question B		b	iii				
a	ii	c	vii				
b	vii	d	vi/ vii				
c	ii	e	vi/ vii				
d	ii/ vii						

Appendix E.3

List of criteria compared across years together with results of statistical tests.

NSD = no significant difference
Statistically significant differences are reported when p<0.01.

Performance of candidates at each band of award in the component criteria of reflective / discursive essays.

Band	Criterion	Statistical significance
Band A	Language	NSD
Band B	Language	NSD
Band C	Language	NSD
Band D	Language	NSD
No award	Language	NSD
Band A	Line of Thought	NSD
Band B	Line of Thought	NSD
Band C	Line of Thought	NSD
Band D	Line of Thought	NSD
No award	Line of Thought	NSD
Band A	Personal Stance	NSD
Band B	Personal Stance	NSD
Band C	Personal Stance	NSD
Band D	Personal Stance	NSD
No award	Personal Stance	NSD
Band A	Other Points of View	NSD
Band B	Other Points of View	NSD
Band C	Other Points of View	NSD
Band D	Other Points of View	NSD
No award	Other Points of View	NSD

Performance of candidates at each band of award in the component criteria of Imaginative writing.

Band A	Language	p = 0.0043
Band B	Language	NSD
Band C	Language	NSD
Band D	Language	NSD
No award	Language	NSD
Band A	Specific Literary Form	NSD
Band B	Specific Literary Form	NSD
Band C	Specific Literary Form	NSD
Band D	Specific Literary Form	NSD
No award	Specific Literary Form	NSD
Band A	Tone/Mood	p = 0.0051
Band B	Tone/Mood	NSD
Band C	Tone/Mood	NSD
Band D	Tone/Mood	NSD
No award	Tone/Mood	NSD
Band A	Thematic Content	NSD
Band B	Thematic Content	NSD
Band C	Thematic Content	NSD
Band D	Thematic Content	NSD
No award	Thematic Content	NSD

Performance of candidates at each band of award in the component criteria of Report writing

Band A	Understanding	NSD
Band B	Understanding	NSD
Band C	Understanding	NSD
Band D	Understanding	NSD
No award	Understanding	NSD
Band A	Selection	NSD
Band B	Selection	NSD
Band C	Selection	NSD
Band D	Selection	NSD
No award	Selection	NSD
Band A	Order and Structure	NSD
Band B	Order and Structure	NSD
Band C	Order and Structure	NSD
Band D	Order and Structure	NSD
No award	Order and Structure	NSD
Band A	Formal Tone and Register	NSD
Band B	Formal Tone and Register	NSD
Band C	Formal Tone and Register	NSD
Band D	Formal Tone and Register	NSD
No award	Formal Tone and Register	NSD
Band A	Language Competence	NSD
Band B	Language Competence	NSD
Band C	Language Competence	NSD
Band D	Language Competence	NSD
No award	Language Competence	NSD

Performance of candidates at each band of award in the component criteria of Literature (Drama)

Band B	Personal Stance	NSD
Band C	Personal Stance	NSD
Band D	Personal Stance	NSD
No award	Personal Stance	NSD
Band A	Relevance/Line of Thought	NSD
Band B	Relevance/Line of Thought	NSD
Band C	Relevance/Line of Thought	NSD
Band D	Relevance/Line of Thought	NSD
No award	Relevance/Line of Thought	NSD
Band A	Knowledge of Text	NSD
Band B	Knowledge of Text	NSD
Band C	Knowledge of Text	NSD
Band D	Knowledge of Text	NSD
No award	Knowledge of Text	NSD
Band A	Textual Analysis Style/Structure	NSD
Band B	Textual Analysis Style/Structure	NSD
Band C	Textual Analysis Style/Structure	NSD
Band D	Textual Analysis Style/Structure	NSD
No award	Textual Analysis Style/Structure	NSD
Band A	Linguistic/Literacy Awareness	NSD
Band B	Linguistic/Literacy Awareness	NSD
Band C	Linguistic/Literacy Awareness	NSD
Band D	Linguistic/Literacy Awareness	NSD
No award	Linguistic/Literacy Awareness	NSD
Band A	Language Competence	NSD
Band B	Language Competence	NSD
Band C	Language Competence	NSD
Band D	Language Competence	NSD
No award	Language Competence	NSD

Performance of candidates at each band of award in the component criteria of Literature (Prose)

Band A	Personal Stance	NSD
Band B	Personal Stance	NSD
Band C	Personal Stance	NSD
Band D	Personal Stance	NSD
No award	Personal Stance	NSD
Band A	Relevance/Line of Thought	NSD
Band B	Relevance/Line of Thought	NSD
Band C	Relevance/Line of Thought	NSD
Band D	Relevance/Line of Thought	NSD
No award	Relevance/Line of Thought	NSD
Band A	Knowledge of Text	NSD
Band B	Knowledge of Text	NSD
Band C	Knowledge of Text	NSD
Band D	Knowledge of Text	NSD
No award	Knowledge of Text	NSD
Band A	Textual Analysis Style/Structure	NSD
Band B	Textual Analysis Style/Structure	NSD
Band C	Textual Analysis Style/Structure	NSD
Band D	Textual Analysis Style/Structure	NSD
No award	Textual Analysis Style/Structure	NSD
Band A	Linguistic/Literacy Awareness	NSD
Band B	Linguistic/Literacy Awareness	NSD
Band C	Linguistic/Literacy Awareness	NSD
Band D	Linguistic/Literacy Awareness	NSD
No award	Linguistic/Literacy Awareness	NSD
Band A	Language Competence	NSD
Band B	Language Competence	NSD
Band C	Language Competence	NSD
Band D	Language Competence	NSD
No award	Language Competence	NSD

Performance of candidates at each band of award in the component criteria of Literature (Poetry)

Band A	Personal Stance	NSD
Band B	Personal Stance	NSD
Band C	Personal Stance	NSD
Band D	Personal Stance	NSD
No award	Personal Stance	NSD
Band A	Relevance/Line of Thought	NSD
Band B	Relevance/Line of Thought	NSD
Band C	Relevance/Line of Thought	NSD
Band D	Relevance/Line of Thought	NSD
No award	Relevance/Line of Thought	NSD
Band A	Knowledge of Text	NSD
Band B	Knowledge of Text	NSD
Band C	Knowledge of Text	NSD
Band D	Knowledge of Text	NSD
No award	Knowledge of Text	NSD
Band A	Textual Analysis Style/Structure	NSD
Band B	Textual Analysis Style/Structure	NSD
Band C	Textual Analysis Style/Structure	NSD
Band D	Textual Analysis Style/Structure	NSD
No award	Textual Analysis Style/Structure	NSD
Band A	Linguistic/Literacy Awareness	NSD
Band B	Linguistic/Literacy Awareness	NSD
Band C	Linguistic/Literacy Awareness	NSD
Band D	Linguistic/Literacy Awareness	NSD
No award	Linguistic/Literacy Awareness	NSD
Band A	Language Competence	NSD
Band B	Language Competence	NSD
Band C	Language Competence	NSD
Band D	Language Competence	NSD
No award	Language Competence	NSD

Performance of candidates at each band of award in the component criteria of Interpretation

Band A	Selecting, Explaining, Summarising Main Concerns & Showing Their Relationships	NSD
Band B	Selecting, Explaining, Summarising Main Concerns & Showing Their Relationships	NSD
Band C	Selecting, Explaining, Summarising Main Concerns & Showing Their Relationships	NSD
Band D	Selecting, Explaining, Summarising Main Concerns & Showing Their Relationships	$p = 0.0078$
No award	Selecting, Explaining, Summarising Main Concerns & Showing Their Relationships	$p = 0.0038$
Band A	Selecting, Explaining Key Words & Phrases & Showing Their Relationships	$p = 0.0000$
Band B	Selecting, Explaining Key Words & Phrases & Showing Their Relationships	$p = 0.0000$
Band C	Selecting, Explaining Key Words & Phrases & Showing Their Relationships	$p = 0.0000$
Band D	Selecting, Explaining Key Words & Phrases & Showing Their Relationships	$p = 0.0000$
No award	Selecting, Explaining Key Words & Phrases & Showing Their Relationships	$p = 0.0000$
Band A	Drawing Inferences From Key Statement(s)	NSD
Band B	Drawing Inferences From Key Statement(s)	NSD
Band C	Drawing Inferences From Key Statement(s)	NSD
Band D	Drawing Inferences From Key Statement(s)	NSD
No award	Drawing Inferences From Key Statement(s)	NSD
Band A	Substantiating Statements, Explanations & Inferences by Reference to Text(s)	$p = 0.0000$
Band B	Substantiating Statements, Explanations & Inferences by Reference to Text(s)	$p = 0.0000$
Band C	Substantiating Statements, Explanations & Inferences by Reference to Text(s)	$p = 0.0000$
Band D	Substantiating Statements, Explanations & Inferences by Reference to Text(s)	$p = 0.0002$
No award	Substantiating Statements, Explanations & Inferences by Reference to Text(s)	$p = 0.0000$

Band A	Making Relevant Comments on Aspects of the Point of View of the Author(s)	NSD
Band B	Making Relevant Comments on Aspects of the Point of View of the Author(s)	p = 0.0050
Band C	Making Relevant Comments on Aspects of the Point of View of the Author(s)	p = 0.0002
Band D	Making Relevant Comments on Aspects of the Point of View of the Author(s)	p = 0.0007
No award	Making Relevant Comments on Aspects of the Point of View of the Author(s)	p = 0.0000
Band A	Substantiating Comments from Evidence in the Text &/or from Their Own Knowledge & Experience	NSD
Band B	Substantiating Comments from Evidence in the Text &/or from Their Own Knowledge & Experience	NSD
Band C	Substantiating Comments from Evidence in the Text &/or from Their Own Knowledge & Experience	NSD
Band D	Substantiating Comments from Evidence in the Text &/or from Their Own Knowledge & Experience	NSD
No award	Substantiating Comments from Evidence in the Text &/or from Their Own Knowledge & Experience	p = 0.0002
Band A	Analysing, Commenting on the Use of Conventions of Genre (incl structure, style, language), & Applying Appropriate Terminology	NSD
Band B	Analysing, Commenting on the Use of Conventions of Genre (incl structure, style, language), & Applying Appropriate Terminology	p = 0.0024
Band C	Analysing, Commenting on the Use of Conventions of Genre (incl structure, style, language), & Applying Appropriate Terminology	p = 0.0030
Band D	Analysing, Commenting on the Use of Conventions of Genre (incl structure, style, language), & Applying Appropriate Terminology	NSD
No award	Analysing, Commenting on the Use of Conventions of Genre (incl structure, style, language), & Applying Appropriate Terminology	p = 0.0000

4 Geography

At an early stage in discussions about what to assess in geography it was decided not to study content which can change from year to year but to consider the fundamental skills of geography. This meant that a complete re-analysis of the examinations from the selected years would need to be undertaken before the scripts could be marked. It was recognised that only someone with extensive knowledge of teaching and assessment in geography could carry out this task. An experienced geographer and former examiner with the SEB was asked to join the working group as a consultant and to explore the feasibility of applying such an analysis to both the unrevised and the revised examinations.

■ Differences between unrevised and revised Higher Grade geography

The unrevised Higher Grade syllabus had three components, geographical ideas through a study of the British Isles, world topics and a practical exercise. The revised syllabus comprises prescribed core studies, selected applications of the core studies and an investigation. The differences are summarised in Table 4.1.

Table 4.1: Unrevised and Revised Higher Grade examinations

	Unrevised examination	Revised examination
Paper I	PART A The physical landscape and countryside PART B Urban settlement	Physical geography Human geography
Format	Eight structured questions. Three questions to be attempted. One each from Parts A and B and one other from any part	Eight short response questions. All questions to be attempted but choice within questions
Paper II	PART A Land and water resources PART B Aspects of social geography PART C Aspects of economic geography	SECTION 1 Rural land resources Rural land degradation River basin management SECTION 2 Urban change and its management European regional inequalities Development and health
Format	Fifteen structured questions five from each aspect Three questions to be attempted. One from Part A, one from Part B or Part C and one other.	Six structured questions one from each application Three questions to be attempted. One each from Sections 1 and 2 and one other Extended writing required
OS maps	OS maps could appear in either part of Paper 1	OS maps appear in Paper 1 only
Coursework	Practical exercise (assessed by the SEB)	Investigation (assessed by the teacher with sample moderation by the SEB)

There are strong similarities between the two courses, but the revised Higher Grade syllabus focuses more on broad systems and their applications to selected area contexts. The practical exercise in the unrevised Higher course and the investigation in the revised Higher course were omitted from the analysis as was the section on economic geography which was an option for all candidates in the unrevised Higher Grade examination and was taken up by few.

Sample size and type

Since the focus for analysis was to be the fundamental skills of geography, opportunities to demonstrate most of these skills were likely to be a part of every geography examination and therefore the need to select two consecutive years in order to ensure coverage of the syllabus was less of a concern than in, for example, mathematics. It was also known that in 1993, only 50 papers had been retained by the Scottish Examination Board and that the number of scripts available at each band would be very small. It was, therefore, decided that scripts from 1987, 1991 (unrevised) and 1994 (revised) should be chosen. Scripts from 100 candidates for each of these years were available and Table 4.2 shows the total number of scripts analysed as part of the geography study.

Table 4.2: Geography sample

Year	Band A	Band B	Band C	Band D	No award
1987	32	16	17	14	18
1991	37	16	15	14	15
1994	33	15	15	15	19

Framework for analysis

Members of the working group spent some considerable time discussing the criteria which should be used to analyse the sample of scripts. In accordance with principles agreed for the study as a whole, the criteria for geography had to

- be recognisable to subject experts as a working synthesis of the subject as taught
- be discrete and observable from the scripts
- achieve a balance between the need to describe detail and the need to work with fundamental aspects of the subject
- be sufficiently robust to accommodate changes in the subject over the years covered by the project.

In order to meet these principles, it was decided that the chosen criteria should not depend on content. There were two main reasons for this decision:

- Each geography examination can include only a sample of the geographic content studied in a Higher Grade course.
- The element of candidate choice within a paper, means that candidates can choose questions based on different content or use their own stock of specific knowledge to illustrate their responses.

A more fundamental set of indicators of ability was required which would provide evidence for the attainment of candidates significant to the study of geography. It was decided to focus on key words within examination questions which signal to the candidates what is expected of them. The list of seven indicators (with, where appropriate, accompanying key words identified in the examination questions) which were finally agreed related to the ability of the candidate to

1. describe... (describe/ what?/ outline/ define/ name/ show)
2. explain...(describe and explain/ account for/ suggest reasons for/ how?/ define and suggest)
3. discuss...(discuss/ comment on/ compare/ assess)
4. demonstrate understanding of geographical processes
5. use/ demonstrate graphic skills effectively
6. interpret and use OS Map Extracts effectively
7. interpret and use effectively other resource materials provided

These indicators were accepted by the working party and later by the group of markers as fundamental for the successful study of geography.

Although the indicators selected as the basis for the analysis were evident in the examination papers for each year, the examinations were neither set nor marked in accordance with these indicators. It was, therefore important to re-analyse each question in each paper to identify the relevant knowledge and skills. Each question could, of course, satisfy more than one indicator. For example in 1987 Paper I Q 7b, the candidate is asked:

With the aid of annotated diagrams, explain the characteristic form and land use pattern of any two of the following

a coastal holiday resort

a New Town

a market town

In this question the candidate is being given the opportunity to

- explain (indicator 2)
- demonstrate understanding of geographical processes (indicator 4)
- demonstrate graphic skills effectively (indicator 5)

In 1994 Paper II Q2b, the candidate is asked:

Study Reference Diagram Q2B and Reference map Q2

With the aid of the graph discuss the impact of rainfall variability on the process of soil erosion.

In this question the candidate is being given the opportunity to

- discuss (indicator 3)
- demonstrate understanding of geographical processes (indicator 4)
- interpret and use effectively other resource materials provided (indicator 7)

The geography consultant was asked to analyse each paper by question or part question in relation to the seven indicators. Table 4.3 shows the number of opportunities offered in each paper for each indicator. Not all opportunities were open to every candidate as there was an element of choice in each paper. This choice was more restricted in the revised examination than in the unrevised examination. This was particularly so in Paper I of the revised Higher Grade where, although there was choice within questions, all questions had to be attempted.

Table 4.3: Opportunities offered in the different papers

	Unrevised				Revised	
	1987		1991		1994	
Indicators	Paper I	Paper II	Paper I	Paper II	Paper I	Paper II
1 Describe	5	4	5	12	7	5
2 Explain	8	12	18	14	12	12
3 Discuss	11	14	9	9	1	5
4 Geographic processes	5	8	5	8	7	5
5 Use graphic skills	1	1	2	0	2	1
6 Interpret OS maps	7	0	9	0	3	0
7 Interpret other resources	7	9	12	10	10	11

The main differences between unrevised and revised Higher Grade examination papers appeared to be in the interpretation of OS maps and in the demand for discussion. Both of these were less in evidence in the revised Higher Grade papers. Comparisons were complicated by the degree of choice given to candidates. For example, in Paper I of the unrevised papers there appeared to be many more opportunities to interpret OS maps. However, the format of these papers was such that candidates had to choose only three questions from a total of eight questions and hence could have avoided all reference to OS maps. In Paper I of 1994 although there were only three instances of the same indicator, all questions had to be answered. Paper I of the revised examination required short responses so the opportunities for discussion were bound to be limited. There appeared to be fewer opportunities to discuss in Paper II of the revised paper than in the unrevised paper but again this was a function of the extent of choice offered in each paper.

■ Levels of attainment

Having selected the indicators to be used, it was decided to use a four point scale for assessing the candidate's level of attainment on each indicator. This scale was written in fairly broad terms to indicate the qualities required to attain each level.

Level 1	Very well expressed, relevant, clear, precise, detailed and showing a high level of understanding
Level 2	Relevant and clear but less well expressed, detailed and perceptive
Level 3	Adequate in terms of relevance, precision and expression
Level 4	Inadequate in terms of relevance, precision and expression.

■ The marking process

It was recognised that only experienced markers who had been involved in setting papers and marking scripts for several years were likely to do justice to the time-consuming and complex task to be undertaken. A small group of five experienced markers who were used to working together were assigned to the task. The external consultant to the project was included in the group of markers.

A half day meeting was set up for markers and the external consultant prepared a session based on candidates' responses to particular questions to raise awareness and to practise re-analysing the papers in terms of the indicators and levels of attainment. Grids were prepared for each question on Papers I and II for each year. These showed clearly which indicators were applicable for each question and part question. Potential difficulties were discussed and agreement reached about how to handle these difficulties. Space was left on the marking sheet for any comments on particular questions or candidates. Each marker marked 20 scripts from each of the three years so that they could all have the experience of working with the indicators and the levels of attainment on both unrevised and revised examination scripts.

■ Analysis of results

As shown in Table 4.3 above, the number of opportunities to demonstrate success on any one indicator differed from Paper I to Paper II and from year to year. For example, in 1987 there were five opportunities to demonstrate success for indicator 1 in Paper I but only one opportunity for indicator 5. For any one candidate the opportunities would be less than those offered as candidates were allowed a degree of choice. Depending on their choice, different candidates would have different number of opportunities within the same paper. The level of success of each candidate was based only on the number of opportunities taken.

Scatter of attainment

A level of attainment was calculated for every candidate on each indicator in each examination paper. For each question or part question, markers assessed the candidate's level of response on a scale of 1 to 4 as shown above. In some cases, where a number of opportunities were offered for the same indicator, the candidate would be a consistent 4 or a consistent 1. For other candidates, the level of attainment was not so consistent. For example, Table 4.4 shows the actual scatter of assessments for four candidates on the same indicator. Each row refers to a different question or part question which provided the opportunity to explain. Because of personal choice or possible lack of time, candidate 001 had six opportunities, candidate 002 had ten opportunities. To determine an overall level of attainment for each candidate on each indicator, a mean score was computed and the mean score was then rounded to the nearest half. This rounding was to avoid spurious differences which were a function of the calculation but which were not warranted by the scale originally used. Mean scores were calculated separately for each paper in the examinations.

*Table 4.4: Indicator 2 - the ability to **explain** (from 1994 Paper I)*

Question number	001	002	003	004
Q1a	n/a	2	3	n/a
Q 1b	3	n/a	n/a	1
Q1c	4	2	2	1
Q2b	n/a	1	2	2
Q3b(i)	n/a	2	2	1
Q3b(ii)	n/a	2	2	1
Q4a	4	2	1	1
Q4c	4	n/a	n/a	n/a
Q5a	3	2	2	1
Q6a	n/a	2	4	1
Q6b	n/a	2	3	2
Q8b	4	1	2	1
Mean score	3.67	1.8	2.3	1.2
Rounded mean scores	3.5	2	2.5	1

The examples shown include an illustration of the scatter of attainments which were sometimes demonstrated in different questions by the same candidate for one indicator. For example, although candidates 001, 002 and 004 are operating fairly consistently at levels 1/2 or 3/4, candidate 003 demonstrates a wide variation in attainment from level 1 to level 4. Most cases where this occurred related to indicator 7 (interpret and use effectively other resource materials provided) and showed candidates who were otherwise operating at levels 1 and 2, demonstrating success at level 4. The reasons for the apparent inconsistency in performance in this particular indicator are discussed later. The inverse rarely occurred, ie candidates who were mainly operating at level 4 rarely showed attainment at level 1.

Based on the rounded mean scores of how candidates performed across a number of opportunities, each candidate was placed on a 7 point scale for each indicator in each paper. For example the attainment levels of the candidates in Table 4.4 were 6, 3, 4 and 1 respectively. The relationship between rounded mean scores and the seven point scale for attainment levels is shown in Table 4.5.

Table 4.5: Converting mean scores to attainment levels

Rounded mean scores	Attainment levels
1	1
1.5	2
2	3
2.5	4
3	5
3.5	6
4	7

Different forms of analysis

Two different ways of analysing the results were used. In the first instance a 'within year' analysis was carried out by a consideration of the performance of candidates in each paper, in each year. Candidates were grouped according to their overall award and the performances between candidates attaining awards of A, B, C and D on each of the seven indicators was calculated.

In the second instance, a 'between year' analysis was used. The information on attainment levels of candidates was reorganised to highlight the differences in performance between years. These differences were calculated for performance in Papers I and II separately and then in the papers combined.

Within year analysis: Success by band level for separate papers

For each overall band awarded, the percentage of candidates who were judged to demonstrate success at each of the seven levels of success was calculated. Two series of graphs are attached as Appendices G.1 and G.2 to show the results.

The graphs in Appendix G.1 show cumulative percentages on the seven point scale and illustrate the differences between the bands on each indicator for 1987, 1991 and 1994. The graphs for the ability to describe in Paper I are copied in Figure 4.1 and the figures upon which the 1994 Paper I graph are based are shown in Table 4.6.

Figure 4.1: Cumulative frequency graphs for the indicator 'Describe'

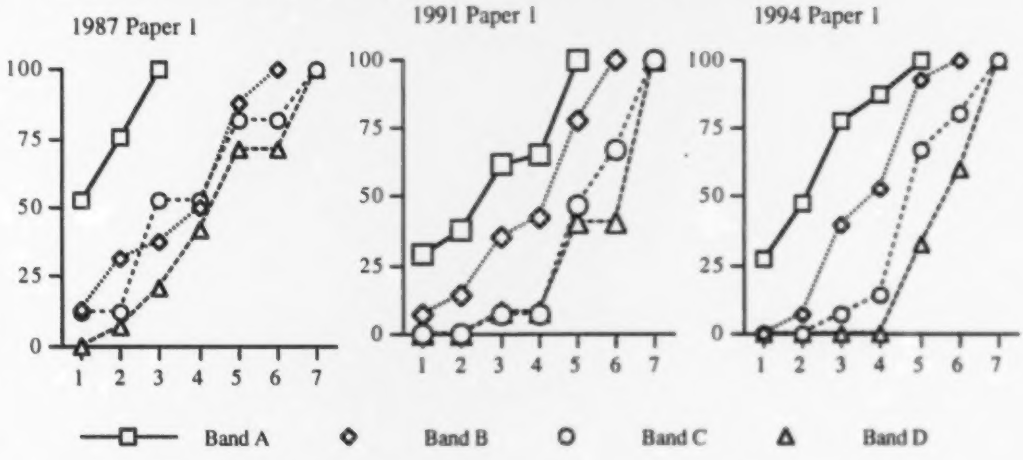


Table 4.6: Percentage of candidates demonstrating attainment in the ability to describe

1994 Paper I	1	2	3	4	5	6	7
Band A	27	21(48)	31(79)	9(88)	12(100)	0	0
Band B	0	7(7)	33(40)	13(53)	40(93)	7(100)	0
Band C	0	0	7	7(14)	53(67)	13(80)	20(100)
Band D	0	0	0	0	33	27(60)	40(100)

The figures in brackets in Table 4.6 show the cumulative percentage frequencies which were used to plot the graphs. The graphs show that, for example, over three quarters of band A candidates in 1994 Paper I attained a level 3 (adequate in terms of relevance, precision and expression) or higher. This compared with 40% of band B candidates, 7% of band C candidates and no band D candidates. In this particular example, no candidates awarded a band D were operating at levels 1 to 4 and no band A candidates were operating at level 6 or 7. In general as the overall band award falls from A to D, the level of ability to describe falls and this is true in both Paper I and Paper II. This trend is repeated in other indicators studied.

The graphs also show that it appears to be easier to differentiate between attainment of candidates at band A and band B than at bands C and D. For example, in the ability to discuss in Paper I of 1991 (unrevised) and 1994 (revised), there is a clear gap between the percentages of band A and band B candidates at levels 1 to 4 (where they might be expected to operate) and much less of a gap between band C and D candidates at levels 4 to 7 (where they might be expected to operate). Examples of very similar performances at band C and band D included attainment on the indicators as follows:

- describe in 1987 Paper II, (Appendix G.1/1)
- explain in 1991 and 1994 Paper II, (Appendix G.1/2)
- discuss in 1991 and 1994 Paper I, (Appendix G.1/3)
- understand geographic processes in 1991 Paper II, and 1994 Paper I, (Appendix G.1/4)

- interpret OS maps in 1991 Paper I, (Appendix G.1/6)
- use resources effectively in 1991 and 1994 Papers I and II (Appendix G.1/7)

Between year analysis

The second set of graphs in Appendix G.2 shows the same information but rearranged so that the differences between the years are highlighted. Again cumulative percentage graphs are used. For example, Table 4.7 matches the first graphs in Appendix G.2/1 which are shown in reduced size as Figure 4.2.

Table 4.7: Band A candidates' ability to ***describe*** *(cumulative percentage)*

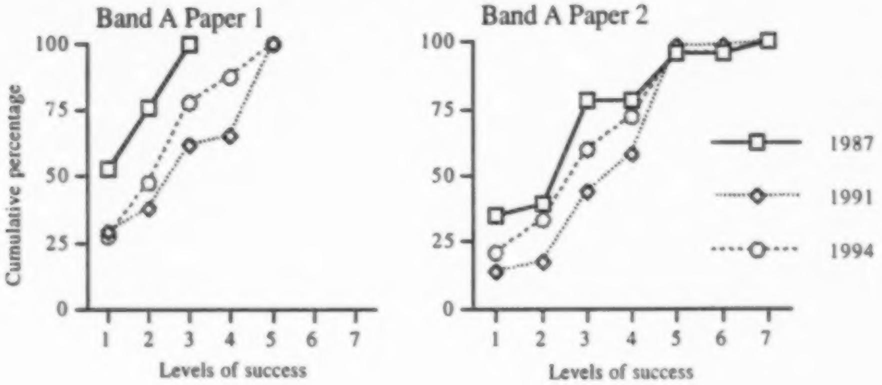
Paper I

Levels	1	2	3	4	5	6	7
1987	53	77	100				
1991	30	39	62	65	100		
1994	27	48	78	87	100		

Paper II

Levels	1	2	3	4	5	6	7
1987	35	39	78	78	96	96	100
1991	14	17	44	58	98	98	100
1994	21	33	60	72	96	96	100

Figure 4.2: Band A candidates' ability to describe (cumulative percentage)



An analysis of significant differences was carried out on Papers I and II separately between the years. The ability to describe was one of the few instances where there was a significant difference. The difference was between 1987 and 1991, in Paper I at band A. In 1987, 77% of band A candidates demonstrated success at level 1 or 2. In 1991, only 39% of candidates were operating at the same levels. Performance in the unrevised papers of 1991 and the revised papers of 1994 were very similar.

The only other significant difference noted when separate papers were considered was in the ability to use graphic skills effectively. This was also in relation to Paper I at band A and the figures are shown in Table 4.8. The equivalent graphs are shown in Appendix G.2/5.

Table 4.8: Band A candidates' ability to **use graphic skills** *(cumulative frequency)*

Paper I

Levels	1	2	3	4	5	6	7
1987	53	53	93	93	93	93	100
1991	24	24	52	52	84	84	100
1994	24	27	57	63	90	90	100

Paper II

Levels	1	2	3	4	5	6	7
1987	17	17	59	59	92	92	100
1991	31	31	56	56	75	75	100
1994	19	23	54	54	81	81	100

Again the difference is between 1987 and 1991 rather than between the unrevised and revised Higher grade examinations.

Mean overall scores

A second analysis was carried out by combining the results from Papers I and II together. Again a mean score was computed for all candidates on each indicator and candidates were grouped by band. Significant differences between the years were then calculated. Table 4.9 summarises the differences. The crosses indicate that when performances in 1987, 1991 and 1994 were compared, there were significant differences between the years.

Table 4.9: Significant differences between the years

Indicators	Band A	Band B	Band C	Band D
1 Describe	x	x		x
2 Explain	x		x	
3 Discuss				
4 Understand geographic processes				
5 Use graphic skills effectively	x	x	x	
6 Use/interpret OS map	x			
7 Use/interpret resource materials	x			

For most of the significant differences at band A the differences were between 1987 and 1991. In each case more candidates, than would have been expected statistically, performed at higher levels in 1987. The differences were not between performance in revised and unrevised examinations.

The effective use of graphic skills was the one indicator where there were differences between performance in unrevised and revised examinations. These differences occurred at bands A, B and C and performance in the effective use of graphic skills was judged to be at a lower level in 1994 than in earlier years. However, the effective use of graphic skills was the indicator where there were fewest opportunities to demonstrate success (see Table 4.3) and therefore the results are likely to be very dependent on question differences. The relevant question in 1987 required an annotated diagram of the characteristics and land use patterns of standard types of settlement. The diagrams requested in 1991 and 1994 were less routine, In 1991, candidates were asked to use an annotated sketch to describe the main features of given areas of an OS map and in 1994, they were asked to use a diagram to help to explain the formation of one river feature selected in response to an OS map question. The question in 1987 was routine compared with those in 1991 and 1994. A wider range of opportunities to demonstrate the effective use of graphic skills would have provided a more reliable measure of candidates' performance in this indicator.

Summary

Methodological issues

Comments from the markers indicated that although the analysis of papers used in this exercise was quite different from the normal marking process, they found the indicators relatively straightforward to apply. The one area where they did face problems was in the assessment of indicator 7 (the ability of candidates to use resources effectively). In some instances candidates used one fact from the diagram as a trigger and wrote extensively about that fact. In other cases it was difficult to tell whether the candidate had made use of the resource or not. If candidates referred to names or figures from the resource then their use of the resource could be assessed. If not, the marker could only use his judgement on the degree to which the resource had been used. In some questions (notably 1991 Paper I Q2a, Q7a i and ii and Q8 i and ii) indicator 7 did not appear to be necessary for a satisfactory answer with the result that candidates who provided an excellent explanation but with little or no reference to what were often unnecessary resources were awarded high scores for indicator 2 but low scores for indicator 7. This was at least in part a function of the marking scheme which had identified these questions as providing opportunities for using resources effectively and markers assessed each candidate on that basis. However, it may be that question setters need to be more precise about the purpose of including resources and where the intention is to assess candidates' ability to interpret resources, this must be made clear to the candidates and be reflected in the marking scheme.

Some concern was also expressed by the markers about the lack of a sense of place or time among candidates. 'There seems to be evidence that candidates have acquired a set of broad generalisations which they apply to any place or area they write about. Apparently all places are identical and interchangeable.' When asked to write about changes in recent years, 'answers referred back as far as the industrial revolution'. The sample of papers was small but nevertheless teachers should perhaps be aware of these concerns expressed by experienced geographers

A second indicator which caused some problems for the analysis was the effective use of graphic skills. Although this had been selected as an indicator of fundamental importance in geography, only a very few instances were noted in each paper. This meant that the results were heavily influenced by question effects ie content and context. Scrutiny of the questions which involved this criteria appeared to indicate that the question in 1987 was of a routine and well practised nature compared to the later years.

Differences in performance between candidates at bands A to D

Comparisons between the performance of candidates in different bands showed the same patterns in individual indicators as in the overall awards ie for each indicator, candidates with higher overall awards performed better than candidates with lower overall awards. Although the indicators used for analysis in this study were not overtly those on which the awards were based, it does suggest that the marking schemes do take account of the same skills. The order of bands appeared to hold in most cases and the differences between band A and band B performances were relatively clear. (See graphs in Appendix G.1) However, on the individual indicators used in this analysis it was sometimes difficult to distinguish between band C and band D candidates.

Range of performance

The range of levels of success demonstrated by candidates of all bands was considerable. Small numbers of band A candidates were assessed as operating at level 4 (Inadequate in terms of relevance, precision and expression) in certain instances of almost all criteria. Conversely one individual band D candidate was judged to be able to 'describe' at level 1 (Very well expressed, relevant, clear, precise, detailed and showing a high level of understanding).

Differences between the years

When significant differences were calculated between the years, most of the differences occurred in Paper I and were between the years 1987 and 1991 rather than between unrevised and revised Higher Grade examinations. One possible explanation for this could be the evolving change in geography over the years. Although most candidates in 1991 were still presented for the unrevised examination, the revised syllabus had been under preparation for some years and perhaps the changes were already being anticipated by geography teachers. On the whole, the differences in performance on the indicators between 1991 (unrevised) and 1994 (revised) were not significant and minor differences sometimes favoured 1991 and sometimes 1994.

The only significant difference between 1987 and 1994 was in the ability to use graphic skills effectively. Candidates appeared to perform at higher levels of attainment in 1987 than in 1994. The influence of question effects on this particular indicator has already been discussed.

Conclusions

In considering the findings of this report, it must be remembered that only a relatively small number of candidates' papers were available for study and for this reason the following findings must be treated with caution.

- The indicators used for the analysis were accepted as fundamental aspects of geography and the performance of candidates on these indicators was largely in line with the overall awards derived from the SEB marking schemes.
- No individual indicators appeared to distinguish clearly between C and D candidates.
- Candidates were not always consistent in their performance on each of the indicators.
- There were no significant differences between the years in the ability of candidates to discuss or in their ability to demonstrate understanding of geographic processes.
- Where there were significant differences, these tended to be between 1987 and 1991 rather than between unrevised and revised examinations.

- When significant differences were studied in relation to separate papers all the differences were in Paper I.
- There was no evidence to suggest that standards of pupil performance between the unrevised and revised examinations had changed.

There are some recommendations for teachers and those involved with the examinations which might be worth considering.

Recommendations

- Teachers might check that candidates know what is expected by the key words which were used as indicators in this exercise and which are used in examination questions.
- The geography panel and principal examiner from the Scottish Examination Board should ensure that a balance of opportunities is offered in relation to fundamental aspects of geography and that sufficient opportunities are provided to get a clear indication of candidates' abilities.
- Question setters should be clear about the purpose of providing resources to accompany questions. Where they are intended to assess candidates' ability to use resources effectively, this must be made clear to candidates and reflected in the marking scheme.

Appendix - Geography

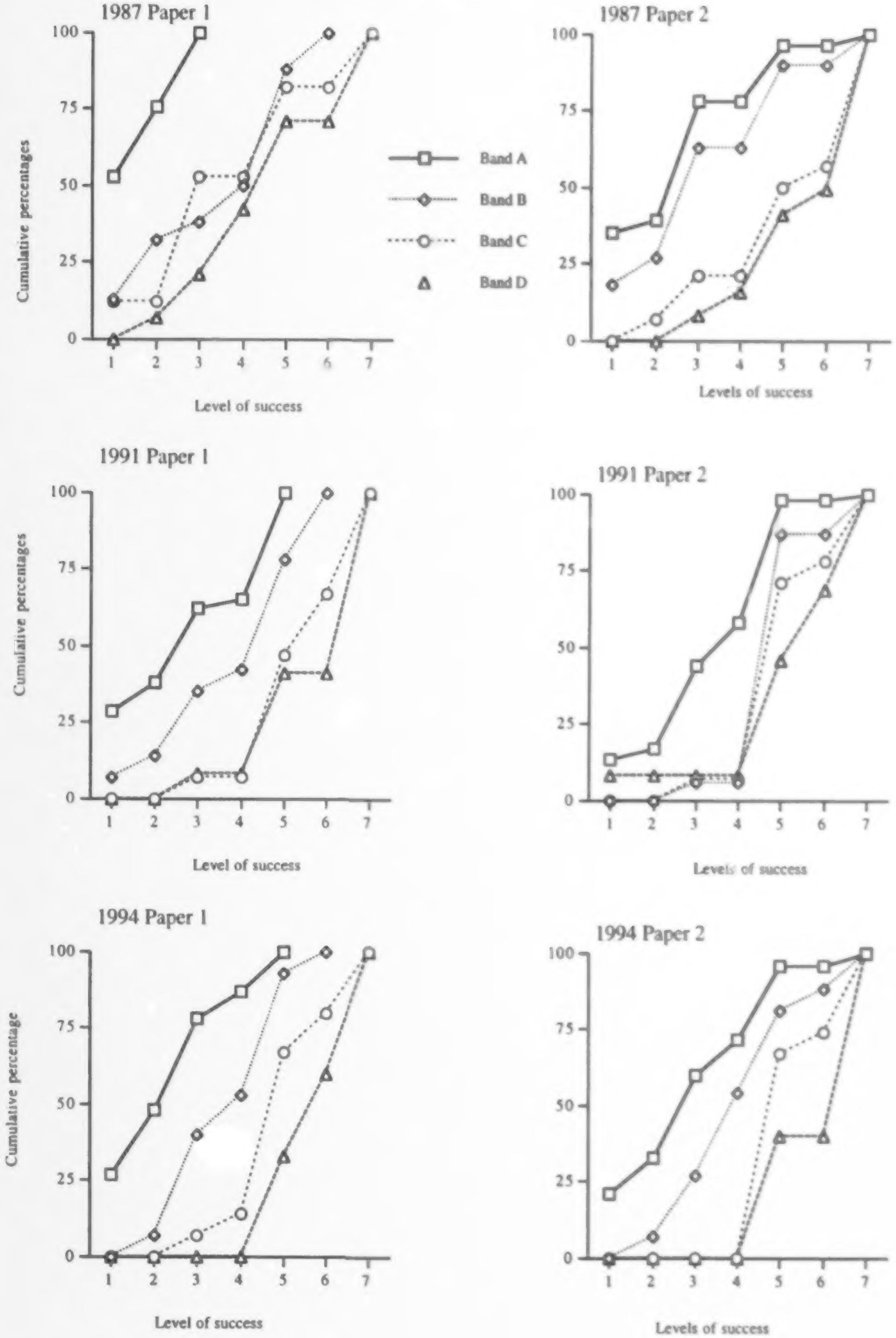
G.1 Cumulative profiles illustrating the differences between the bands

G.2 Cumulative profiles illustrating the differences between the years

List of working group and markers

Appendix G.1/1

The ability to describe



1987 Paper 1

1987 Paper 2

1991 Paper 1

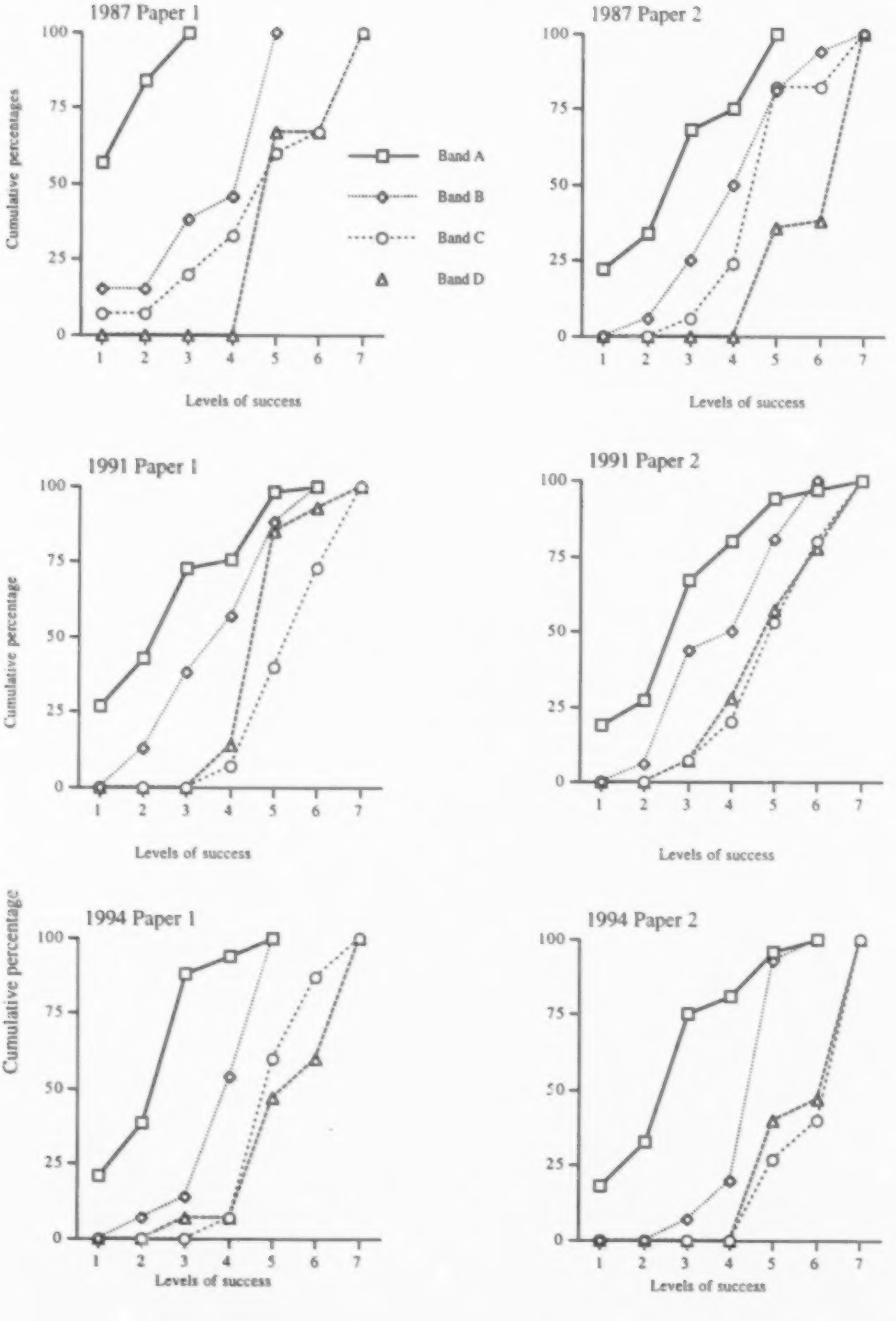
1991 Paper 2

1994 Paper 1

1994 Paper 2

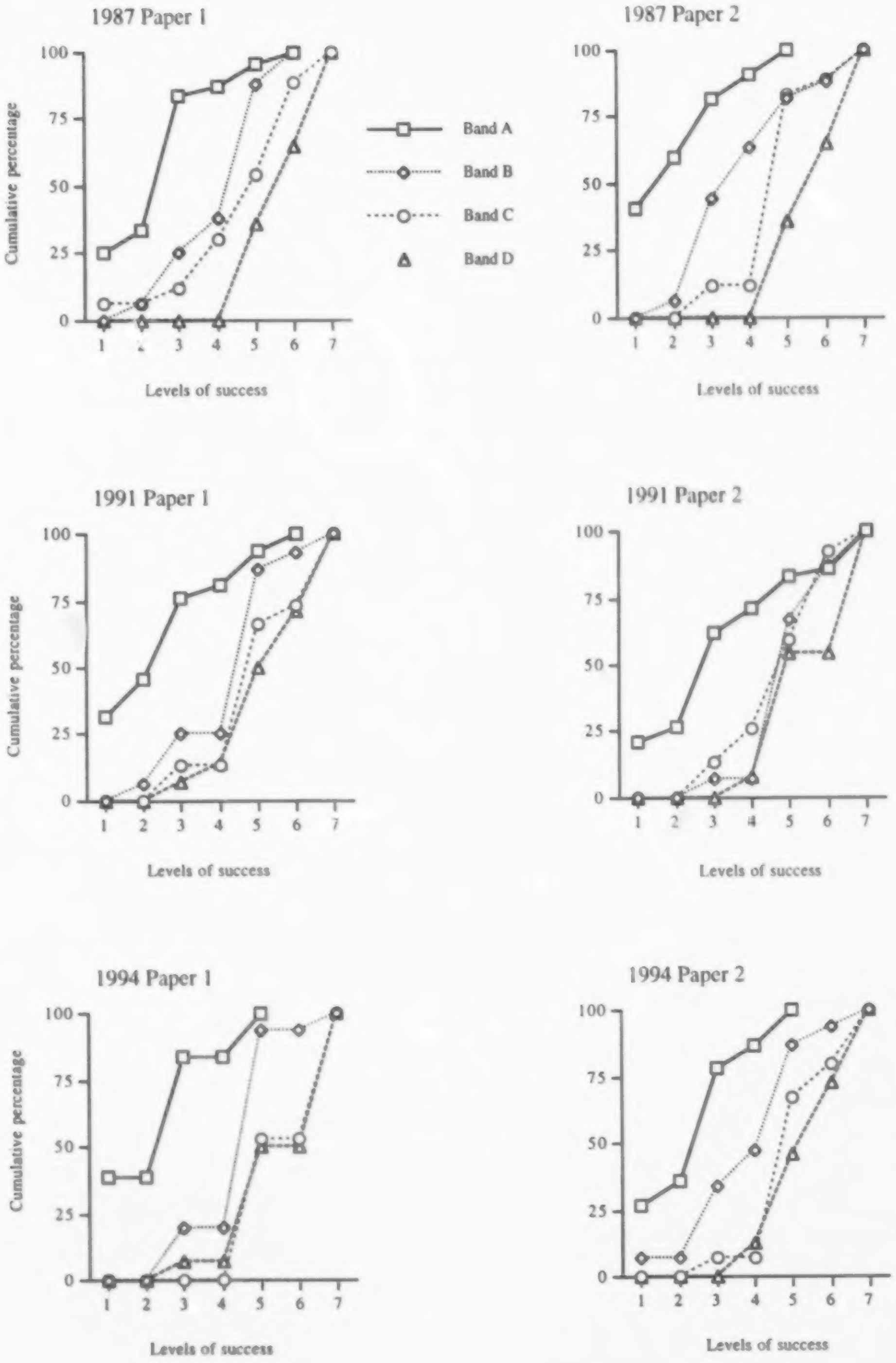
Appendix G.1/2

The ability to explain



Appendix G.1/3

The ability to discuss


1987 Paper 1

1987 Paper 2

1991 Paper 1

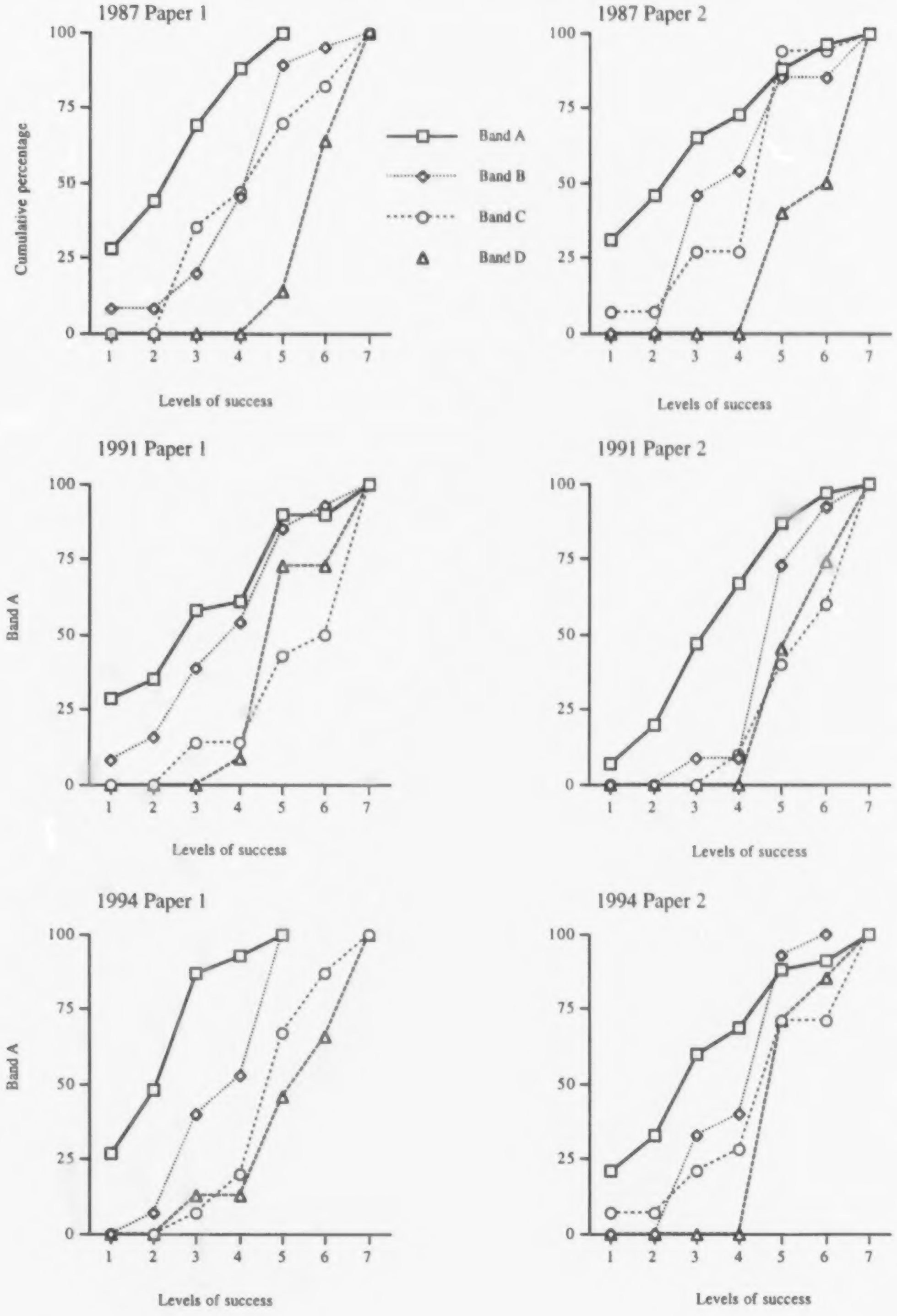
1991 Paper 2

1994 Paper 1

1994 Paper 2

Appendix G.1/4

The ability to understand geographic processes

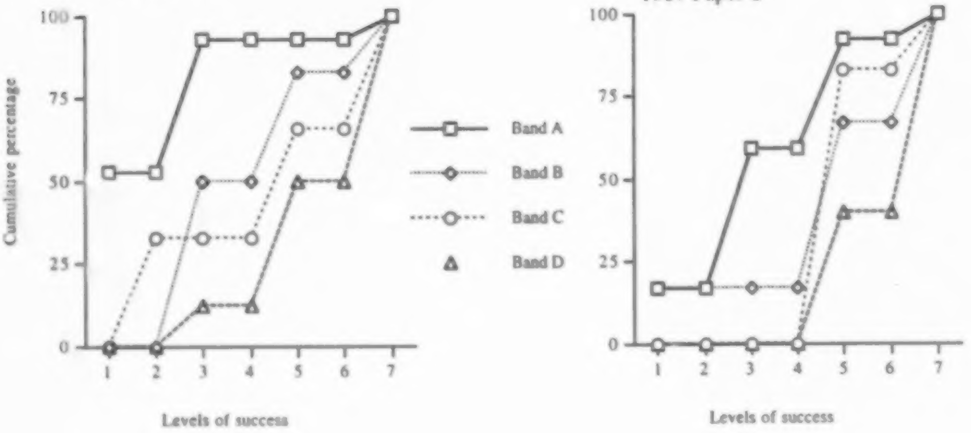


Appendix G.1/5

The ability to use graphic skills

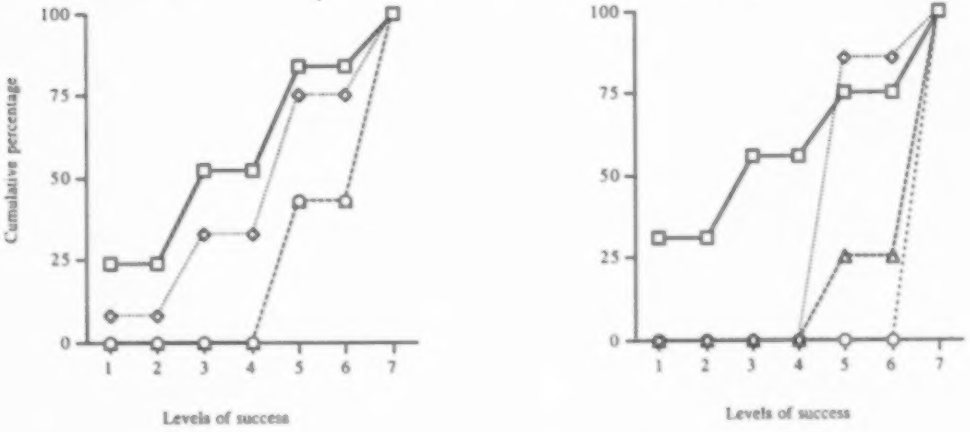
1987 Paper 1

1987 Paper 2



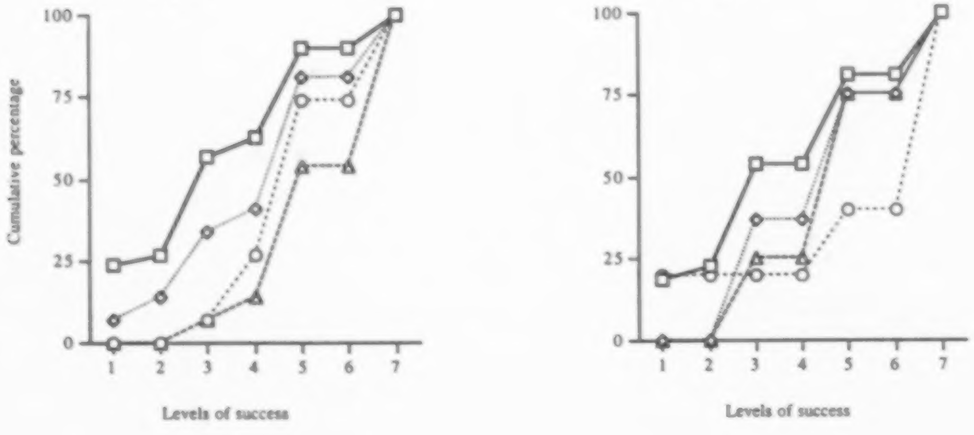
1991 Paper 1

1991 Paper 2



1994 Paper 1

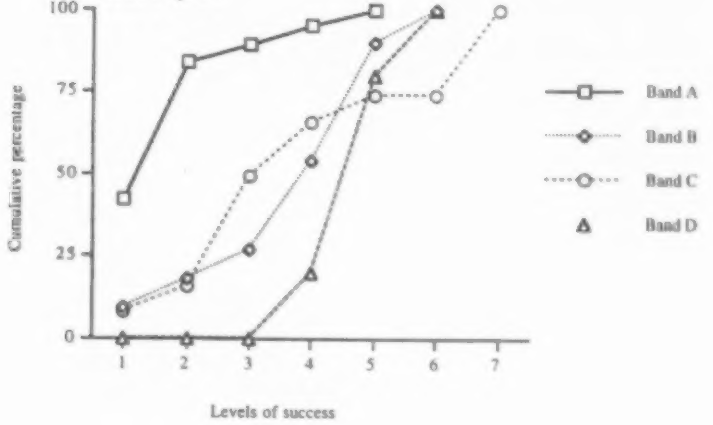
1994 Paper 2



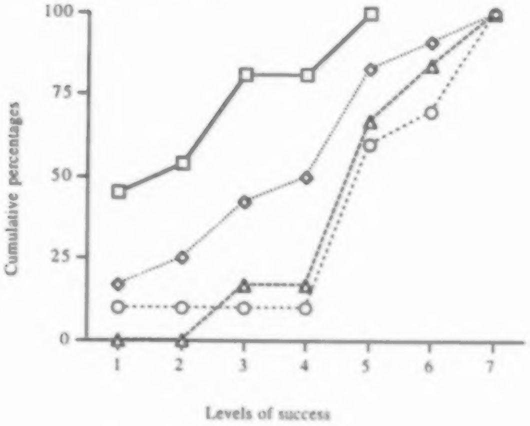
Appendix G.1/6

Use of OS maps (in Paper 1 only)

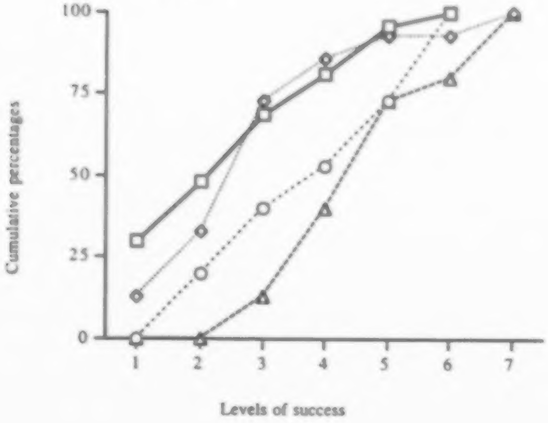
1987 Paper 1



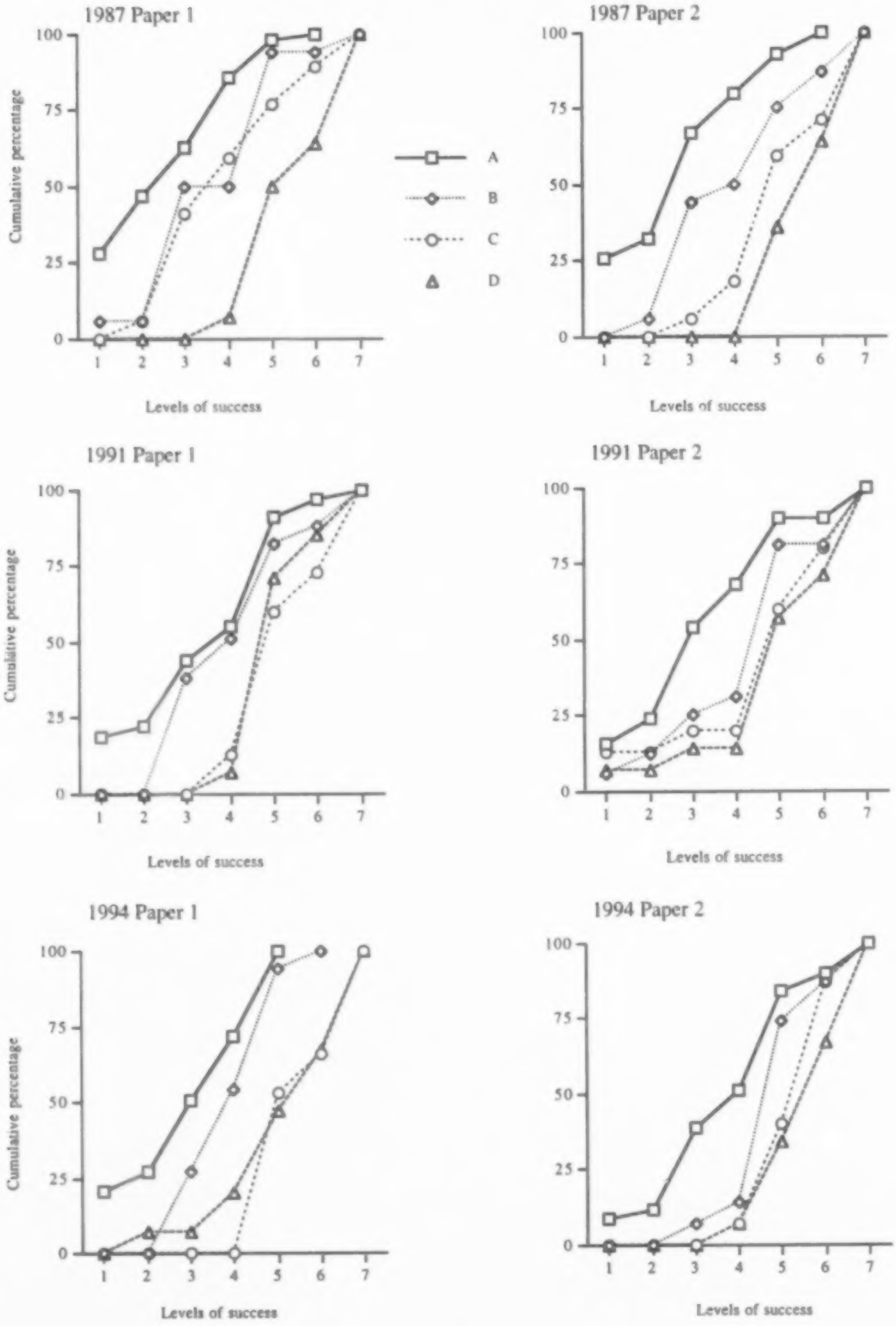
1991 Paper 1



1994 Paper 1



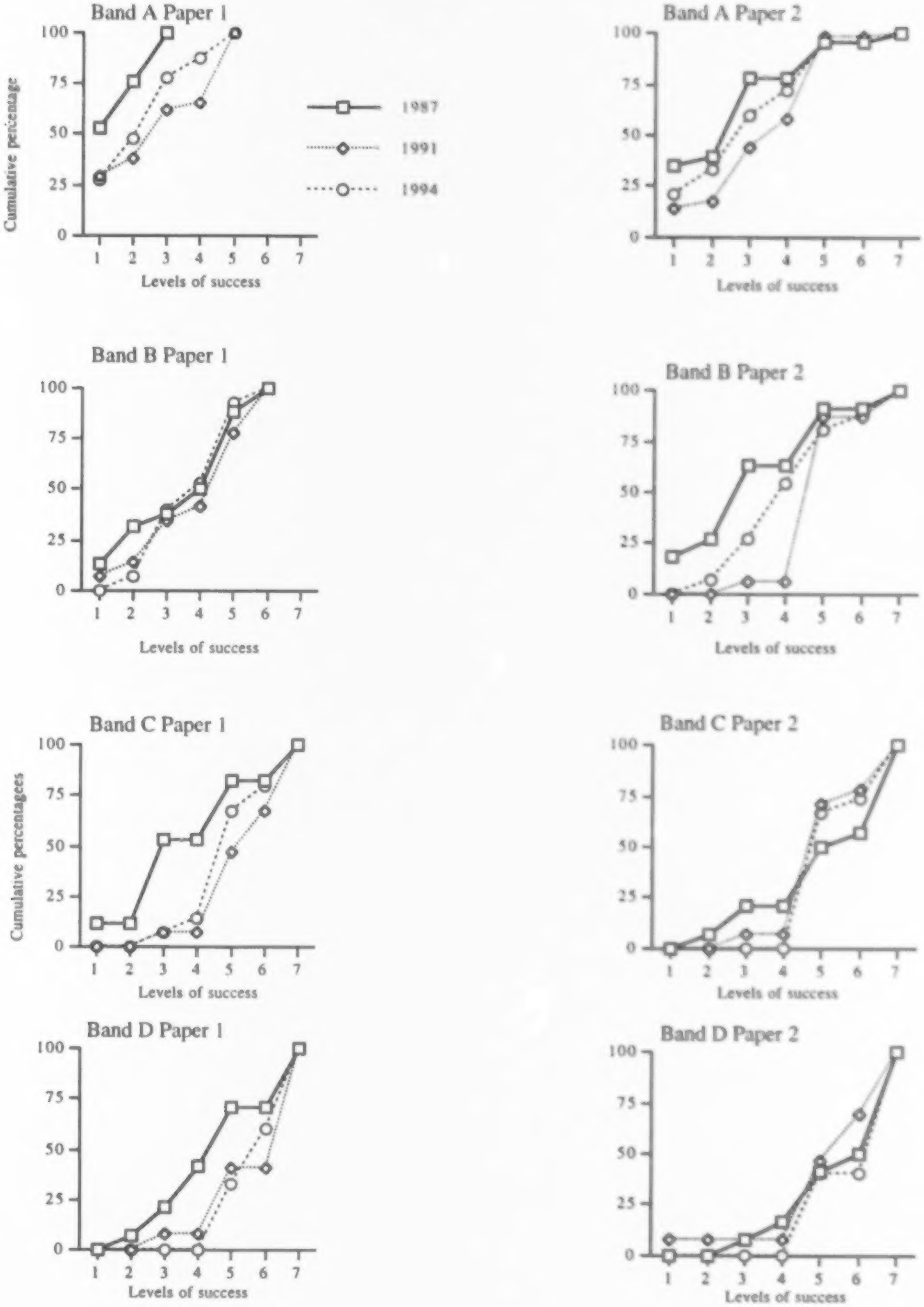
Appendix G.1/7



Use resources effectively

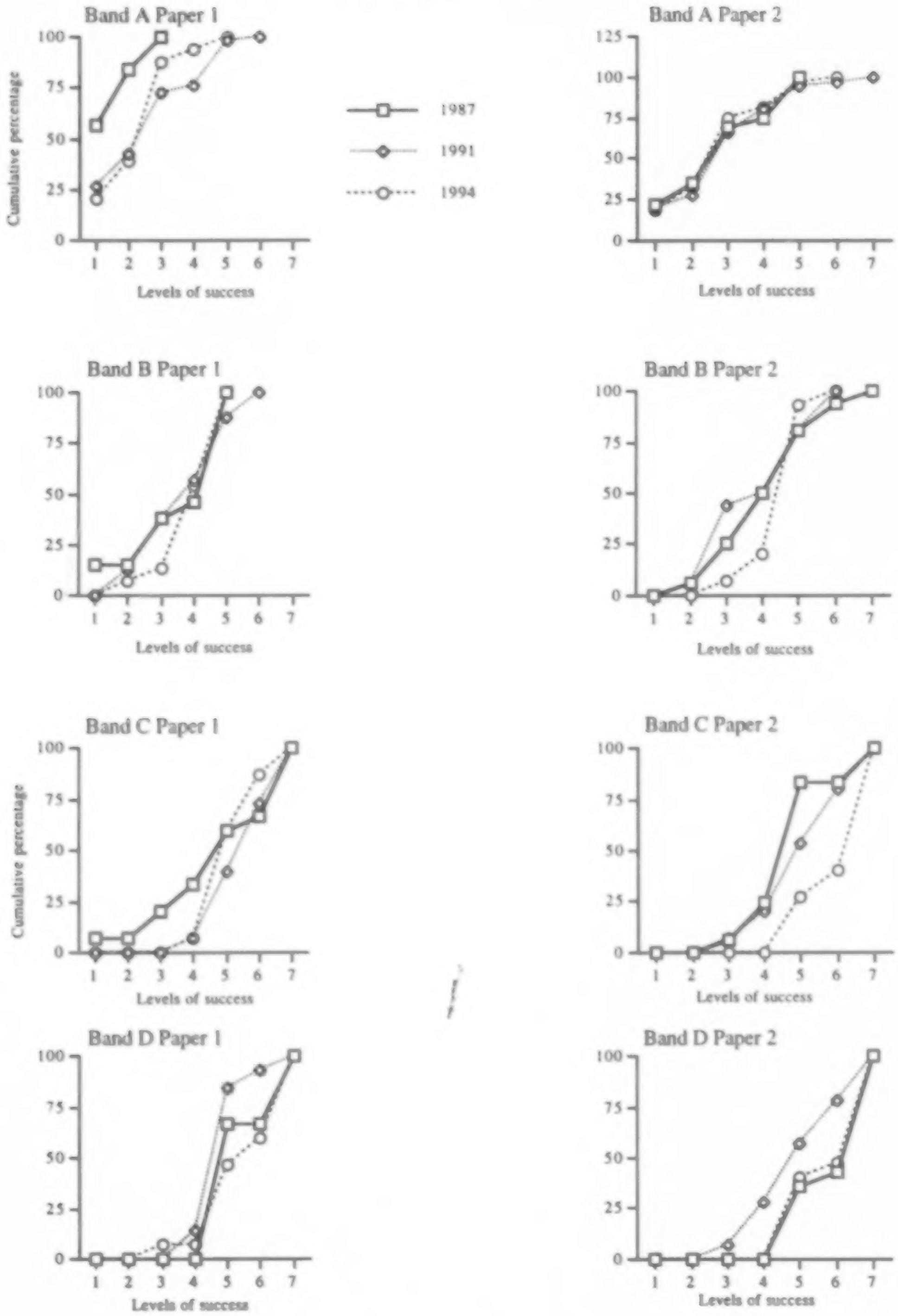
Appendix G.2/1

Ability to describe



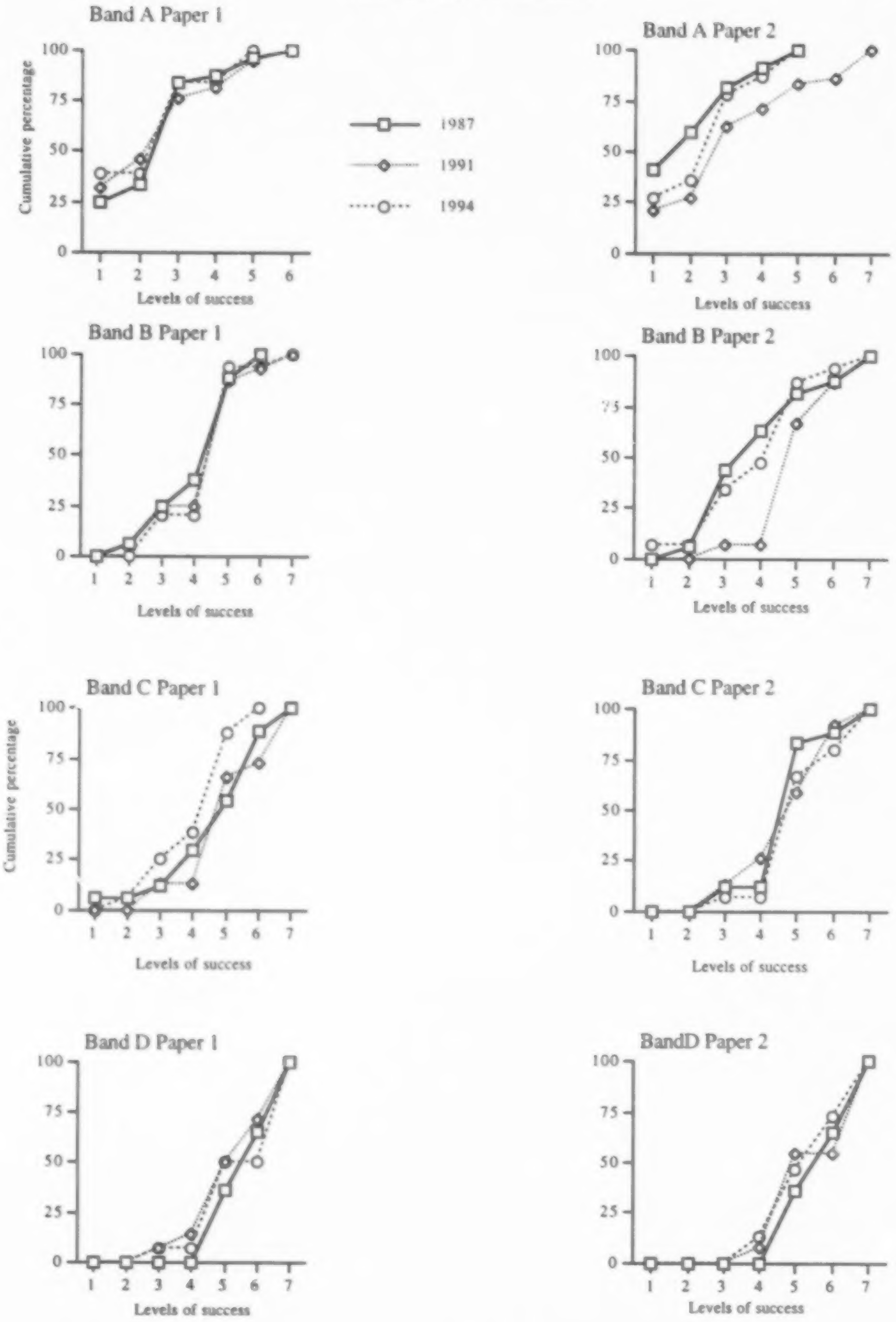
Appendix G.2/2

The ability to explain



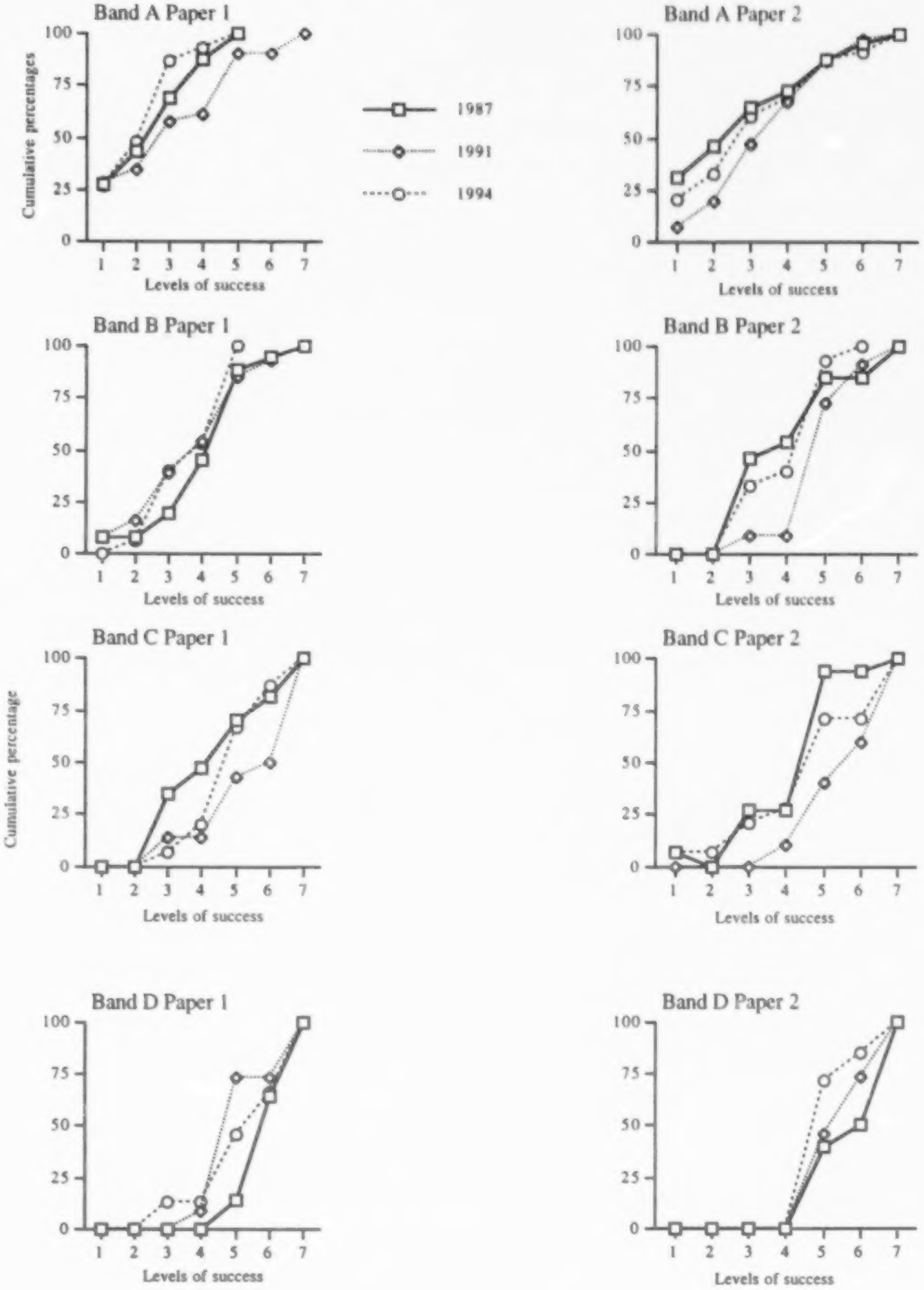
Appendix G.2/3

The ability to discuss



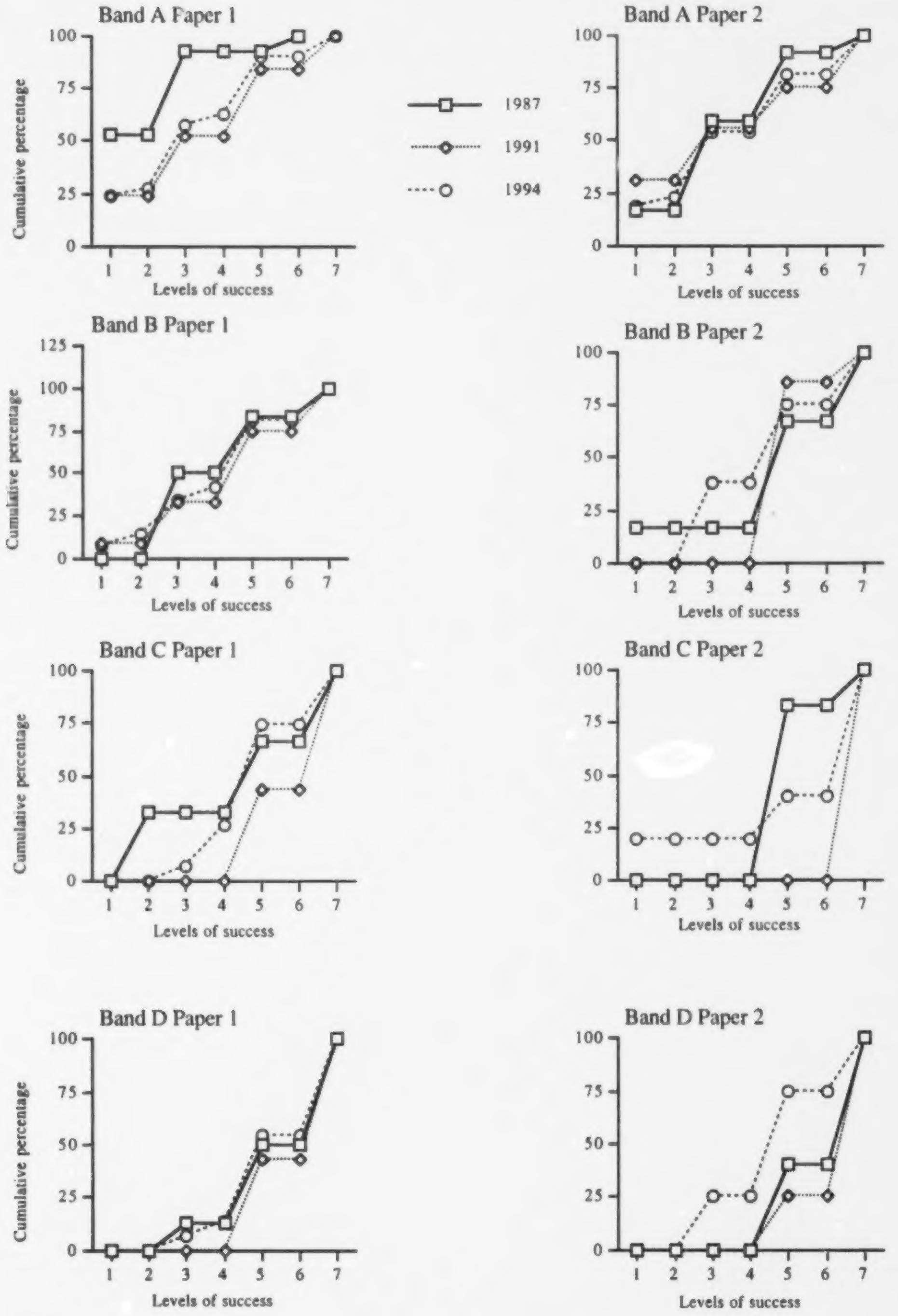
Appendix G.2/4

Understanding geographic processes



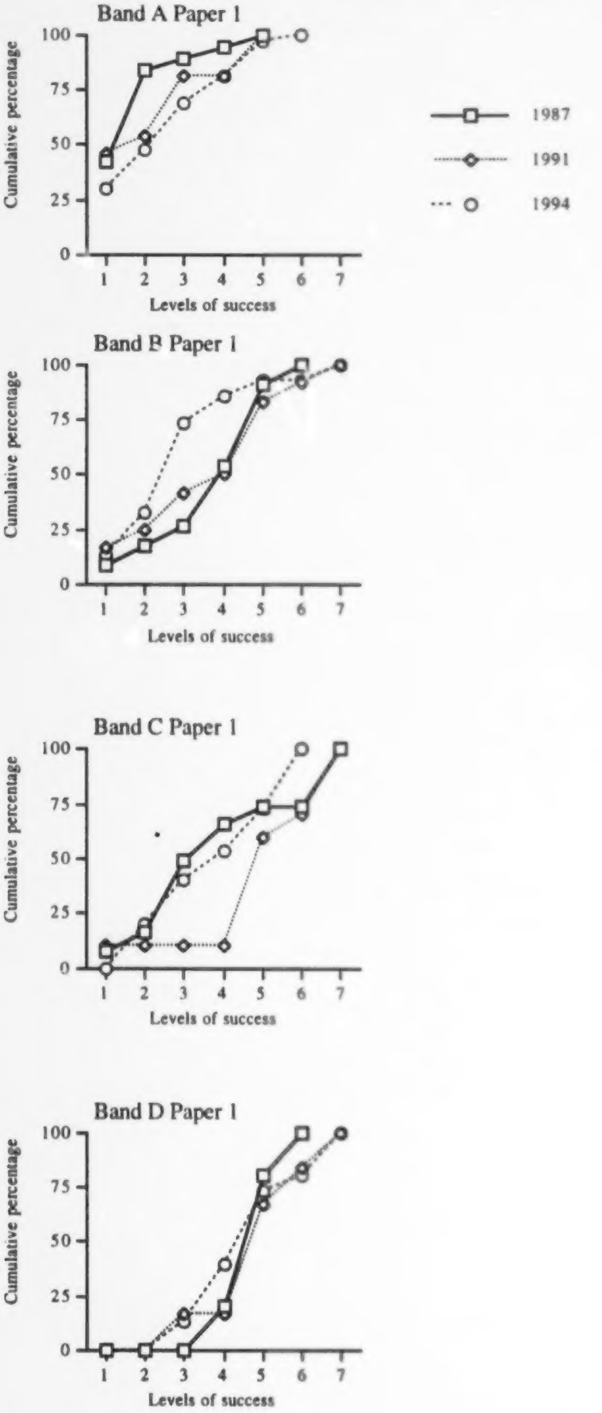
Appendix G.2/5

Effective use of graphic skills



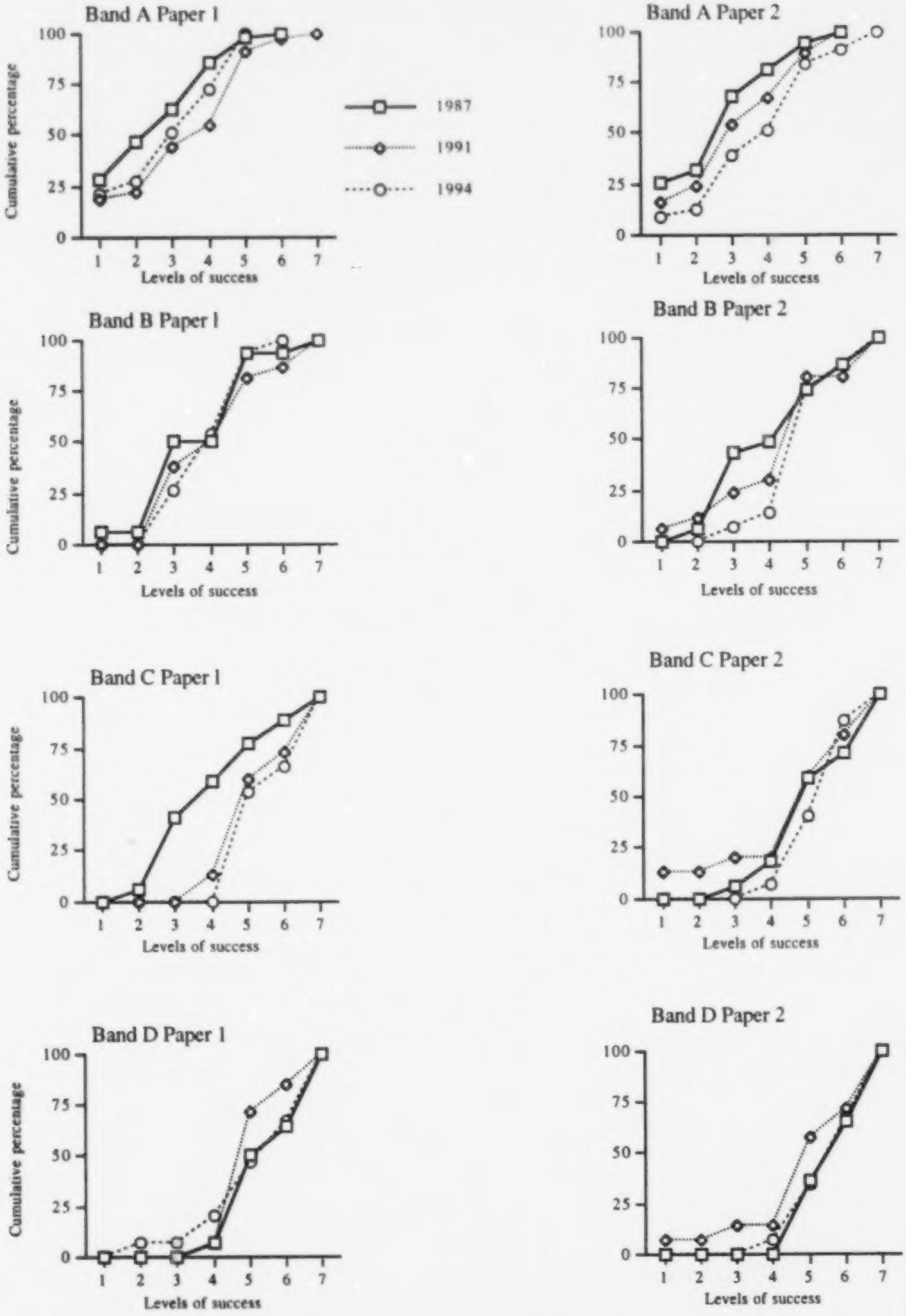
Appendix G.2/6

Interpretation of OS maps



Appendix G.2/7

Use of resources



5 Mathematics

During early meetings of the mathematics working group, participants spent some considerable time discussing the best ways of comparing performance across unrevised and revised Higher Grade examinations. It was recognised that it would not be easy to prepare a list of criteria which were common to both sets of papers and that only someone with extensive knowledge of both syllabuses would realistically be able to undertake this task. An experienced former examiner with the SEB was asked to join the working group as a consultant and to prepare a preliminary paper setting out possible comparisons and highlighting some of the potential problems.

Differences between unrevised and revised Higher Grade mathematics

Comparisons between pupils' performances on unrevised and revised Higher Grade examinations in mathematics are made difficult by a number of important differences between the respective syllabuses and the examinations. These differences arise from the content of the syllabuses, from the approach to mathematics and from the nature of the assessment instruments.

Content

The unrevised syllabus contained mathematical content which is no longer required in the revised syllabus. For example, the earlier syllabus required candidates to

- present routine proofs
- know the properties of arithmetic and geometric series
- know and use expressions for the sum and product of the roots of a quadratic equation
- know a series of definitions and conditions associated with matrices
- associate matrices with geometric transformations
- recall formulae for sums and products of sines and cosines
- determine derivatives from first principles
- know and use the rule for change of base in logarithms

None of these are required for the revised Higher Grade examination.

There is also some content which is appropriate only to the revised Higher Grade syllabus. This includes

- application of recurrence relations
- increased emphasis on graphicacy

The significant pruning of content in the revised Higher Grade syllabus allows scope for more in-depth problem solving, work set in unfamiliar and 'real life' contexts and for an investigative component.

Approach

In the revised examination many of the questions are set in context and demand the application of mathematical procedures. They are in contrast to questions in the unrevised paper which were more dependent on the ability to perform routine algorithms. For example, the ability to use information about the centre and radius of a circle to give the equation of a circle occurs in both syllabuses but the questions in the examination papers differ considerably. The more straightforward question is in Paper II of 1988 and the more complex question is in Paper I of 1994.

1988 Paper II

Given that A and B are the points (4, -1) and (-2, 7) respectively, state the centre and radius of the circle on AB as diameter and write down its equation.

1994 Paper I

The straight line $y = x$ cuts the circle $x^2 + y^2 - 6x - 2y - 24 = 0$ at A and B.

Find the co-ordinates of A and B

Find the equation of the circle which has AB as diameter.

(Diagram included)

In the first case candidates are led through the solution being told the value of the co-ordinates and being asked to find the centre and the radius of the circle as a step towards writing down the equation of the circle. In the second case, they have to find the co-ordinates and are expected to know that they also need to find the co-ordinates of the centre and the radius before they can find the equation of the circle.

In addition, the revised Higher Grade examinations contain questions which introduce unfamiliar mathematics. For example, in 1993 Paper II, Question 4 is based on the eigenvalues of a matrix. An example of how to find the eigenvalue of one matrix is given and candidates are then asked to use this information to calculate eigenvalues and related information for other matrices. Since matrices are no longer included in the revised Higher Grade syllabus, the terminology and format are quite unfamiliar to candidates.

Setting mathematical problems in a 'real-life' context constitutes another difference in approach between the unrevised and the revised Higher Grade examination. Questions which involved substituting limits in a definite integral were provided in 1988 Paper II, Question 5 and 1994 Paper II. Question 10. Note that both questions appear in Paper II of the respective years. In 1988 it was a straightforward question.

Evaluate: $\int^2 (3x - 5)^7 \, dx$

In 1994 the question involved the parabolic cross section of a cargo boat and candidates were asked to find the area of the cross section. The candidates had to interpret the problem, consider what was relevant, select a strategy, implement a strategy and communicate the result.

Problem solving is another area where it is difficult to make comparisons. In the revised examination candidates are expected to work their way through a problem consisting of several steps with little or no support. In the unrevised paper, the candidates are led though a problem step by step making it difficult to evaluate their independent problem solving ability.

Format of examination

The unrevised Higher Grade examination consisted of a fixed response test of 40 questions each worth 2 marks and a second paper of approximately 12 questions worth 100 marks. The revised Higher Grade examination consists of a first paper of approximately 20 short response questions worth 80 marks and a second paper of approximately 10 extended response questions worth 100 marks. An externally assessed investigation contributes 10% towards the total mark.

Basis for comparison

In attempting to come to an agreement about what to compare, the difference in format of Paper I in the unrevised examination and Paper I of the revised examination created difficulties. Paper I of the unrevised examination was a multiple choice paper and Paper I of the revised examination is open response. Comparing responses which depend on selecting the correct answer and responses which are completely open may create differences which are solely the function of the different formats. It was therefore decided to concentrate on Paper II in the unrevised examination and Papers I and II in the revised examination, all of which are open response papers. The investigative component of the revised examination had no parallel in the unrevised examination and was also omitted from the analysis.

Several different ways of making comparisons were discussed. Decisions had to be made about both the scale and type of criteria to be used in the analysis. As a first step, the various papers for each year were analysed question by question to determine what was being assessed. Other SEB documents such as the Arrangements and final marking instructions were also studied. The criteria were then listed and numbered. They ranged from specific content such as the ability to solve inequations to more general skills such the ability to handle unfamiliar situations. A total of 180 criteria were eventually identified and the re-marking of scripts was carried out in relation to these criteria. Appendix M.1 gives the complete list. Not all of these criteria were evident in both the unrevised and the revised examination.

Sample size and type

In mathematics, the examination for each year can only assess part of the total Higher Grade syllabus. In order to have as wide a coverage of the syllabus as possible, it was decided to study papers from two consecutive years and the years selected were 1987/88 (unrevised) and 1993/94 (revised). Scripts from 100 candidates for each of these years were available and these were distributed across the bands as described in the introduction (Table 1.1). The total number of scripts analysed for mathematics is shown in Table 5.1.

Table 5.1 Sample of mathematics scripts by band

Year	Band A	Band B	Band C	Band D
1987	35	13	15	15
1988	35	18	14	15
1993	34	16	16	16
1994	34	16	16	16

It was decided not to include those scripts which received no award and a few scripts which were used during the trial marking process were also omitted.

■ The marking process

It was recognised that only experienced markers who had been involved in setting and marking scripts for several years were likely to do justice to a time-consuming and complex task. Past experience of the SEB suggested that, because of the detailed nature of the marking instructions, there was little disagreement among mathematics markers in marking candidates' scripts and that double marking was not necessary. A small group of five markers who were used to working together were assigned to the task and these included the external consultant and the Examination Officer from the SEB.

A half day meeting was set up for markers and a proforma prepared to help them in their task. Each question paper was analysed question by question for all possible criteria and a recording sheet was provided which enabled the markers to indicate if the criteria was achieved, not achieved or not attempted. A number of papers was selected for markers to practise marking and come to joint decisions about any difficulties.

As explained earlier, there is a considerable difference in the nature of the questions in the unrevised and revised Higher Grade papers. In the unrevised papers the questions were largely straightforward and dependent on knowledge of well known algorithms and routines. In the revised papers the questions were often set in context and placed greater emphasis on independent problem solving abilities. The ability of candidates to attain a certain criterion in the revised examination could, therefore, be masked by their inability to work through the problem. It was, considered essential that candidates should only be marked wrong if there was some evidence that they had attempted the criterion.

At the markers' meeting the following decision was made. If a candidate worked through the first part of a three part question and failed to complete the second and third part, then he/she would be marked as failing in the criteria for the second part (because he/she had information to carry out the second part but did not do so) but would be marked as a non attempt for the third part (because he/she did not have the information to carry out the third part). For example, Q5 in 1994 Paper II was as follows:

a) Express $3\sin x^\circ - \cos x^\circ$ in the form $k\sin x(x - a)^\circ$, where $k > 0$ and $0 \le a \le 90$.

b) Hence find algebraically the values of x between 0 and 180 for which $3\sin x^\circ - \cos x^\circ = \sqrt{5}$.

c) Find the range of values of x between 0 and 180 for which $3\sin x^\circ - \cos x^\circ \le \sqrt{5}$.

A candidate who had completed part a) but neither part b) nor part c), would be marked as incorrect on the second part (because the information was available) but non attempt on the third part (because the information was not available).

This distinction between failure and non attempt was considered to be crucial and much of the training session was devoted to looking at examples where this occurred.

Over a period of about three weeks each marker marked papers either from 1987 and 1993 or 1988 and 1994. This enabled all markers to have the experience of using the criteria on both an unrevised and a revised examination. The external consultant marked a smaller number of papers from each year so that he could act as arbitrator should any problems arise during the marking process. Markers were also asked to note for each question, as appropriate, comments on two general aspects of mathematics

- presentation
- logical layout of working

These notes were used to assist markers in assigning an overall grade of 1 to 4 for each candidate in each of these aspects.

■ Analysis of results

The criteria identified in the examination scripts were grouped in four different ways and an analysis carried out on each group. The four ways are briefly summarised below.

- *Non-comparable criteria:* Two criteria were selected as significant examples of aspects of mathematics which cannot be compared because they do not appear in both syllabuses. One was specific to the unrevised syllabus and one to the revised syllabus.
- *Individual comparable criteria:* A total of 180 individual criteria were identified as being assessed in the four selected years. A sub-set of these appeared in both the unrevised and the revised examinations and were judged to be of comparable difficulty.
- *Basic grouped criteria:* Criteria were grouped together according to topic. The basic groups included those instances of criteria which were judged to be straightforward.
- *All grouped criteria:* All instances of criteria within each topic formed the final group.

Attempt rates

For each method of analysis, the first step was to find the attempt rate for each criterion. Particularly at bands B, C and D the number of papers available was small (see Table 5.1). It was therefore decided to check the numbers attempting each criterion in order to eliminate from the analysis those which were being attempted by less than 50% of band C candidates. There were two reasons for this: results from a handful of candidates would have been difficult to interpret and the revised Higher Grade examination is set so that only 65% of the paper is written at the band C level. This means that only a few band C and band D candidates would be expected to be successful in the more difficult criteria making the problem of small numbers even more acute. In practice only about 10% of the criteria were omitted. In a number of cases, repetition of the same criteria in different questions meant that where one instance of a criterion was omitted because of a low attempt rate in one question, a second instance of the same criterion would be retained in another question.

Attainment on non-comparable criteria

The changes in content between the unrevised and revised examinations meant that certain criteria could only be studied in relation to one of the examinations. Although this study is primarily concerned with comparing performance between the examinations, two criteria have been selected, one from each of the unrevised and revised examinations, to provide examples of what candidates attained in criteria specific to one syllabus only. From the unrevised syllabus the criterion selected was

- C169: the ability to appreciate the requirements of a proof

From the revised syllabus, the criterion selected was

- C171: the ability to cope with unfamiliar situations

Table 5.2 shows the percentages of candidates demonstrating attainment for each criterion.

Table 5.2: Percentage of candidates demonstrating attainment

	Band A	Band B	Band C	Band D
C169 (unrevised)	100	97	86	87
C171 (revised)	100	100	100	75

The results show that in the years in which these criteria were taught and assessed, at least three quarters of all candidates demonstrated attainment.

Attainment on individual comparable criteria

The second analysis looked at the comparison between candidates' performance in unrevised and revised examinations on individual criteria. This involved calculating the percentages of candidates at different bands who demonstrated attainment.

Several different questions shared instances of the same criterion, and a score was calculated for each candidate based on the number of correct responses out of the number of opportunities. A cut off score was set to define attainment for each criterion. Because the emphasis of the analysis was on what candidates could do, this cut off score was set deliberately low.

- If there was only one opportunity, it had to be correct
- If there were two opportunities, one out of two had to be correct
- For three or more opportunities candidates had to have 60% or more correct.

It was then possible to calculate the number (or percentage) of candidates at each band who demonstrated attainment for each individual criterion. Since two consecutive years had been chosen to allow better coverage of the syllabus and since the main concern was to study performance across unrevised and revised syllabuses, the number of candidates demonstrating attainment was combined for each pair of consecutive years - 1987/1988 and 1993/94. This aggregation provided larger numbers of candidates for the analysis than would have been possible by studying the years separately.

Good comparisons

Of the 180 criteria which were identified, approximately 60 of these appeared in both the unrevised and revised examinations. These criteria are marked with a single asterisk in Appendix M.1. Even where two questions or part questions from different years have been selected as assessing the same criterion, the questions will not be exactly comparable. The difference in the content of particular questions can make comparisons difficult. The external consultant prepared a list of criteria which on the basis of his professional judgement and experience were considered to be good comparisons. The full working group had the opportunity to consider this list and to amend it as necessary. As an example of a good comparison, Q9a and Q13a from 1988 Paper II and 1993 Paper I respectively were identified as assessing the ability of candidates to 'find formula for $f(g(x))$' - Criterion 006 in Appendix M.1.

Question 9a (1988 PII)	Functions $f(x) = x+a$ and $g(x) = bx^2$, where a and b are rational constants, defined on the set of real numbers. Find expressions for $(f \circ g)(x)$ and $(g \circ f)(x)$.
Question 13a (1993 PI)	$f(x) = 2x - 1$, $g(x) = 3 - 2x$ Find a formula for $k(x)$ where $k(x) = f(g(x))$.

An example of a 'poor' comparison assessing the same criterion is Q9a from 1988 as above and Q19 from 1994 Paper I.

Question 19 (1994 PI)	A function f is defined on the set of real numbers by $f(x) = x/(1-x)$ $(x \neq 1)$ Find, in its simplest form, an expression $f(f(x))$.

This question is of a different order of difficulty partly because of the complexity of dealing with an algebraic quotient and partly because of the use of a single function. Although in practice there is no difference between calculating $f(g(x))$ and $f(f(x))$ candidates will be much more familiar with a calculation which involves two different functions than one which involves a function of the same function.

For this analysis dealing with the percentage of candidates demonstrating attainment on individual criteria, good comparisons only were used. These comparisons are marked with a double asterisk on the list in Appendix M.1.

Presentation and discussion of results

A series of graphs were prepared for each of the individual criteria which were deemed to have been attempted by more than 50% of band C candidates and which were judged to be comparable. There were 42 instances of comparable criteria in this list and they covered the major areas of mathematics including algebra, geometry, vectors, trigonometry, differential and integral calculus. Table 5.3 shows this list.

Table 5.3 List of comparable criteria

Algebra	Geometry	Vectors	Trigonometry	Diff. Calculus	Int. Calculus
Find formula for f(g(x))	Know radius and tangent are perpendicular	Find 3D vector components from co-ordinates	Given trig ratio find angle	Differentiate a polynomial	Know to integrate f'(x) to find f(x)
Identify remaining quadratic in synthetic division	Use distance formula with co-ordinates	Evaluate scalar product of column vectors	Given 1st quad. angle find related angles	Differentiate +ve fractional index	Introduce constant of integration
Square out a bracket	Find mid-point of line joining two points	Know how to find the angle between vectors	Use expansion of sin 2x	Use chain rule in trig differentiation for multiple angles	Find constant of integration by substitution
Substitute linear equation in quadratic	Substitute co-ordinates in gradient formula	Use scalar product = 0, implies perpendicularity	Use acosx + bsinx = rcos (x-c)	Use chain rule in trig differentiation for power	Deal with coeff of x in integration of $(px + q)^n$
Simplify to quadratic	Use gradients of perpendicular lines result		Know how to find r in rcos (x-c)	Know to differentiate for gradient	Substitute limits in definite integral
Factorise quadratic	Substitute point and gradient in equation of line		Know how to find angle c in rcos (x-c)	Know to differentiate for stationary points	
Identify roots from factors	Find intersection of two lines		Know to use rcos (x-c) to solve equations	Find y values of roots of derivatives	
Perform algebraic simplifications	Use centre, radius to find equation of circle		Solve equation of the form rcos (x-c) = k	Find nature of stationary points	
				Equate dy/dx to zero for stationary points	

The statements above and in Appendix M.1 are written in a concise summary form for illustrative purposes only. Discussions between all those involved in analysing and marking the scripts ensured that there was a shared understanding of each statement.

A set of bar graphs showing the percentage of candidates at each band demonstrating attainment on each criterion is included in Appendix M.2. Bars for unrevised and revised examinations are set alongside each other. At band A, there are very high percentages of candidates being successful in both examinations and therefore few differences occur. Some differences occur at the lower bands. For example, in criterion C028 (the ability to square out a bracket) a higher percentage of band B candidates demonstrated attainment in the revised examination. Conversely, in criterion C129 (know how to find angle c in rcos(x-c)) a higher percentage of band B candidates demonstrated attainment in the unrevised examination. The direction of these differences is almost evenly distributed between unrevised and revised results. At bands C and D, there was a larger number of differences but these too were fairly evenly distributed between unrevised and revised results.

As a means of summarising the information contained in the bar graphs, Table 5.4 provides a count of the differences at each band which occurred in all 42 criteria.

Table 5.4: Number of differences

Band	A	B	C	D
Equal percentages of successful candidates	17	13	6	1
Larger percentage successful in unrevised exam	11	15	19	19
Larger percentage successful in revised exam	14	14	17	22

Equal percentages of band A candidates demonstrated attainment in 17 of the 42 criteria over the unrevised and revised examinations. In 11 criteria, more band A candidates were successful in the unrevised examinations. Conversely, in 14 criteria, more band A candidates were successful in the revised examination. For the other bands the trend was similar. In about half of the criteria which showed a difference, more pupils demonstrated attainment in the unrevised examination. In the other half more pupils demonstrated attainment in the revised examination.

The evidence appears to indicate that where good comparisons can be made between the performance of candidates in the unrevised and revised examinations, standards of performance have been maintained.

Attainment on basic grouped criteria

The third analysis depended not on individual criteria but on grouped criteria. It was recognised that although the bar graphs for individual criteria are important and useful for understanding attainment at the level of specific aspects of mathematics it would also be useful if these criteria could be grouped together in some way to give a more rounded picture of mathematics.

At present, the revised Higher Grade syllabus uses grade descriptions to help to define the content of the syllabus. Some examples of these grade descriptions are

- Given the graph of a function f(x) the candidate can sketch the graphs of simple related functions such as -f(x), f(x) + 2, 3f(x)
- The candidate knows how to determine the area bounded by the curve y = f(x), the x-axis and the lines x = a and x = b.
- The candidate can change expressions into an equivalent form by correctly applying the laws of indices.

The possibility of matching different groups of criteria to the grade descriptions was considered and rejected. First, these descriptions are provided for band C only with additional notes to differentiate between band C and higher bands. Second, it would have been difficult although not impossible to fit the grade descriptions to the criteria in the unrevised examination. More importantly, with the developments in Higher Still, the grade descriptions of the current revised

Higher Grade syllabus may not be directly applicable in future years. It was decided, therefore, to return to topic descriptions based on content and to group the criteria accordingly. Four groups of criteria were identified- Algebra, Geometry, Trigonometry and Calculus. The basic groups for these topics included all instances of relevant criteria in either the unrevised or the revised examinations which were judged to be relatively straightforward. For example, in 1987 Paper II, Question 12b was judged to be a difficult instance of using trigonometric formula for the sum of sines and cosines:

Q.12 (b) Prove for $0 < x < \pi/4$,

$(\sin 5x + \sin x)/(\sin 3x - \sin x) = 2\cos 2x + 1$

In 1994 Paper I, Question 6 was judged to be a straightforward instance of the same criterion.

Q.6 Find algebraically the exact value of $\sin J^\circ + \sin (J + 120)^\circ + \cos (J + 150)^\circ$

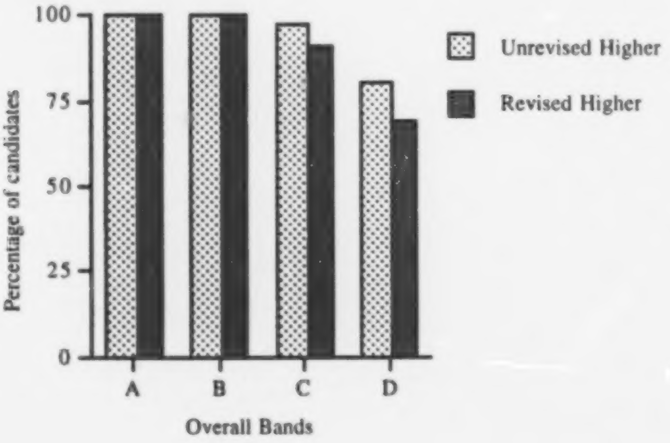
The judgement on which instances of the various criterion were straightforward was first made by the external consultant and later confirmed by the full working group for mathematics.

Each topic consisted of several criteria and in many cases there was more than one instance of a criterion in each year. It was therefore necessary to define a cut off score as a level of attainment. In each case the level of attainment was set at 70% ie for candidates to demonstrate attainment, they had to show acceptable responses in 70% or over of the instances of criteria attempted within each group.

Basic algebra

A total of 44 instances of basic algebraic manipulations were identified within the unrevised examination papers and 68 within the revised papers. These included such criteria as squaring out a bracket and factorising quadratic expressions. The percentage of candidates at each band who demonstrated attainment was calculated separately for unrevised and revised examinations and Figure 5.1 shows the results.

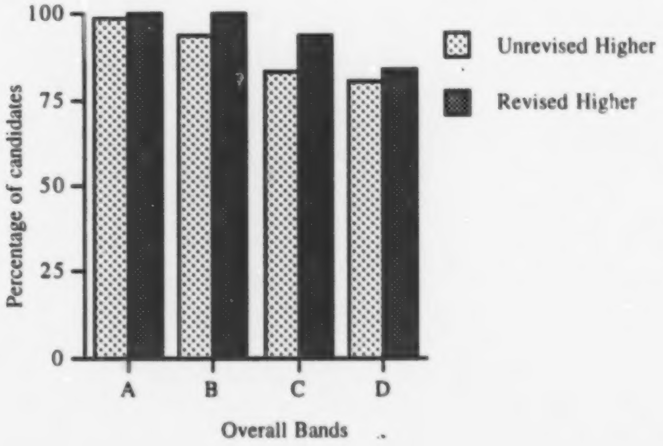
Figure 5.1 Basic algebra manipulations



All candidates at bands A and B, demonstrated attainment in both the unrevised and revised examinations. At the lower grades, slightly fewer candidates demonstrated attainment in the revised examinations than in the unrevised examinations. However, even at band D in the unrevised examination, over two thirds of candidates demonstrated success.

Basic geometry
For basic geometry, those criteria relating to straight line geometry such as substituting co-ordinates in the gradient formula and finding the intersection of a straight line with the y axis were chosen. A total of 17 criteria in the unrevised examinations and 35 criteria in the revised examinations were identified within this grouping. The percentages of candidates demonstrating attainment at each band were calculated and Figure 5.2 shows the results.

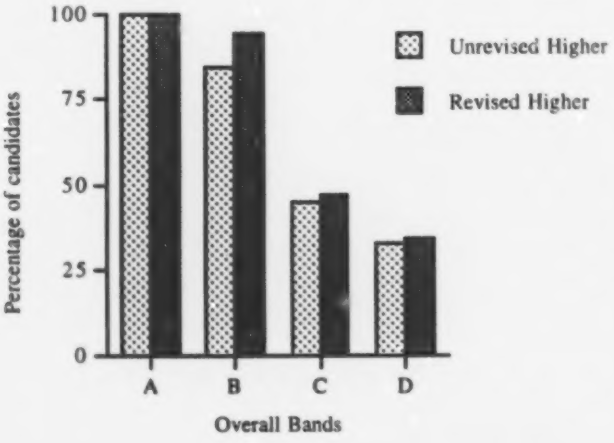
Figure 5.2 Basic geometry



The percentages of candidates at each band were largely similar with slightly more candidates demonstrating attainment in the revised examinations.

Basic trigonometry
For basic trigonometry, 19 instances of criteria were identified in the unrevised papers and 30 in the revised papers. These criteria related to trigonometric ratios and compound angle formulae. The results are shown in Figure 5.3.

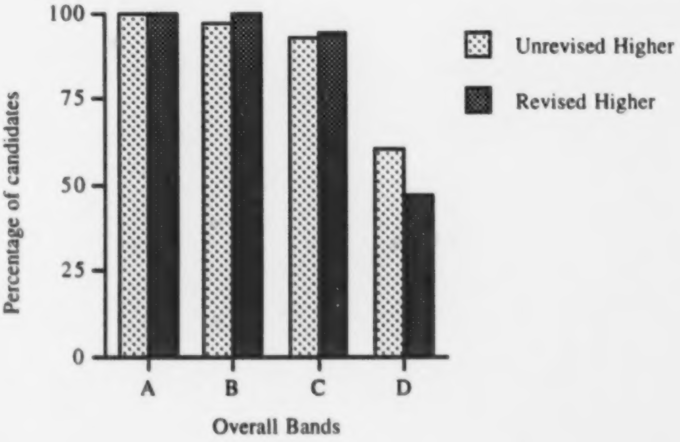
Figure 5.3: Basic trigonometry



The percentages of candidates showing evidence of attainment at each band were very similar. Slightly more candidates demonstrated attainment in the revised examinations. This was the only group where the percentage of candidates demonstrating attainment at bands C and D fell below 50%.

Basic calculus
A total of 35 instances of criteria related to calculus was identified in the unrevised papers and 43 instances in the revised papers. Figure 5.4 shows that there were very few differences except at band D, where a larger percentage of candidates showed evidence of attainment in the unrevised than the revised Higher. In both examinations the band D candidates performed substantially less well than other bands.

Figure 5.4: Basic calculus



Standards in basic grouped criteria
Where groups of relatively straightforward criteria were clustered together under topic headings, there were no significant differences between performance in revised and unrevised examinations. Standards relating to attainment in straightforward instances of algebra, geometry, trigonometry and calculus were unchanged.

Attainment on groups comprising all instances of relevant criteria
The fourth analysis looked in detail at the level of attainment of all candidates at each of the various bands. Once again grouped criteria based on mathematical topics were used. In this case all instances of those criteria which were relevant to each group were included both those which were judged to be relatively straightforward and those which were judged to be more difficult. The same topic groups were used but geometry was extended beyond straight line geometry and calculus was sub-divided into differential and integral calculus.

Since each group contained instances of criteria which had been judged to be difficult, the attempt rate of candidates at each band was likely to have an effect on the level of attainment. As a first step in the analysis the attempt rate for each candidate was calculated. For all candidates at each band, the total number of correct responses and the number of correct and incorrect responses in each topic area were computed.

In accordance with the intention not to penalise candidates for non attempts, calculations for attainment were based on the number of correct responses divided by the total number of all responses (both correct and incorrect). Instances of criteria where no attempt was made were omitted for individual candidates.

Attempt rates (percent) for grouped criteria

For each area of mathematics, the number of opportunities to attempt the criteria is given. Table 5.5 shows the percentage of criteria attempted at each band. For example, in the group of criteria related to algebra, 95% of the 78 instances of criteria in the unrevised examinations (1987/88) were attempted by band A candidates and 97% of the 123 instances of criteria in the revised examinations (1993/94) were attempted by band A candidates.

Table 5.5: Opportunities and percentage attempt rates for grouped criteria

Algebra	**Unrevised**	**Revised**
Number of opportunities	78	123
Band A	95	97
Band B	89	83
Band C	81	68
Band D	69	54

Geometry	**Unrevised**	**Revised**
Number of opportunities	50	94
Band A	91	95
Band B	78	91
Band C	75	89
Band D	72	81

Trigonometry	**Unrevised**	**Revised**
Number of opportunities	33	62
Band A	90	91
Band B	76	79
Band C	52	64
Band D	40	46

Differential calculus	**Unrevised**	**Revised**
Number of opportunities	27	51
Band A	99	93
Band B	99	84
Band C	100	67
Band D	96	56

Integral calculus	**Unrevised**	**Revised**
Number of opportunities	11	19
Band A	99	99
Band B	97	92
Band C	96	90
Band D	76	70

The trend is for the attempt rate to fall with the band level. In most cases, the criteria in the unrevised examination appear to be more accessible to the lower attaining candidates than those in the revised examination. This could be explained by the different nature of the questions in both examinations. Examples provided towards the beginning of this section show that in the unrevised examination, the questions are direct and free standing. In the revised paper questions are often set in a problem solving context where candidates have to decide the steps for themselves. The difference in attempt rate is non existent or negligible in the case of band A candidates demonstrating their ability to do well in mathematics no matter what form the questions take.

Trigonometry was the only area where the attempt rate by band seemed to fall more sharply in the unrevised examination than in the revised examination. As part of the changes introduced in the revised examination, candidates were provided with a page of formulae including trigonometric formulae for the sums and products of sines and cosines. In the unrevised examinations these formulae had to be recalled or developed by the candidates themselves. This could be at least one explanation for the low attempt rates in the unrevised papers.

It was assumed that non attempts would have more effect on attainment in revised examination questions than in unrevised examination and a check for this effect was carried out. Attainment in this case was calculated on the basis of correct responses divided by the total number of opportunities ie non attempts were counted as failure. As an example the figures for differential calculus are shown in Table 5.6.

Table 5.6: Attainment when non attempts are included as failure

Differential Calculus	**Unrevised**	**Revised**
Band A	95	90
Band B	91	72
Band C	83	54
Band D	72	37

In Table 5.6, the percentages of instances of relevant criteria being attained at different bands in the unrevised examination fell from 95% at band A to 72% at band D. In the revised examinations, the fall was much greater, from 90% to 37%. These figures confirm that lower attaining candidates in the revised examinations can access fewer opportunities than in the unrevised examinations. In order to eliminate this effect and to establish what candidates can do, the analysis reported below focused only on those instances of criteria which had been attempted.

Attainment rates based on attempts only

Criteria which were not attempted by individual candidates were omitted from the analysis and the calculation was based on the number of correct responses divided by the number of attempts. For example, in the group of criteria related to algebra, 97% of the instances of criteria in the unrevised examinations which were attempted by band A candidates were completed successfully. Table 5.7 show the results for each topic area.

Table 5.7: Correct responses as a percentage of those criteria attempted

Algebra	**Unrevised**	**Revised**
Band A	97	96
Band B	90	87
Band C	85	76
Band D	71	66

Geometry	**Unrevised**	**Revised**
Band A	94	98
Band B	84	92
Band C	79	86
Band D	67	80

Trigonometry	**Unrevised**	**Revised**
Band A	96	96
Band B	82	84
Band C	67	69
Band D	60	52

Differential calculus	**Unrevised**	**Revised**
Band A	97	96
Band B	91	86
Band C	83	79
Band D	75	65

Integral calculus	**Unrevised**	**Revised**
Band A	98	92
Band B	94	81
Band C	86	71
Band D	74	61

The actual percentages of attainment show that Higher Grade candidates who obtain a pass (bands A - C) can do more than two thirds of all the instances of criteria in these topic areas which they attempt. The trend follows the pattern of bands. As the band falls, the attainment rate falls. Except in geometry, the trend is more pronounced in the revised examination than in the unrevised examination and raises the question of whether candidates at the lower grades are being less successful than their counterparts in the unrevised examinations. The differences were tested statistically.

Significant differences

Statistically, geometry was the only topic where there were significant differences between the unrevised and revised Higher Grade performance at all bands. More correct responses were noted in the revised papers than were statistically expected. Other differences noted were in integral calculus for bands A, B and C, in differential calculus for bands B and D and in algebra for band C. In these cases the number of correct responses was higher than statistically expected in the unrevised examinations. Table 5.8 summarises the differences and indicates the direction. For example, 'unrevised' means that the number of correct responses was higher than statistically expected in the unrevised examinations.

Table 5.8: Significant differences

Topic	Band A	Band B	Band C	Band D
Algebra			Unrevised	
Geometry	Revised	Revised	Revised	Revised
Trigonometry				
Differential calculus		Unrevised		Unrevised
Integral calculus	Unrevised	Unrevised	Unrevised	

One possible reason for differences which indicate more success in the unrevised papers may relate to the comparability of the criteria. The instances of criteria which were grouped together to form the topic areas, included both those instances which were judged to be straightforward and all other instances assessing the same competence. Not all of these other instances of criteria were of comparable difficulty and examples of good and poor comparisons have been discussed above. It may be that the poor comparisons were not evenly distributed across the years and that the more difficult instances were mainly in the revised papers. The fact that there were no significant differences in the criteria associated with calculus when only straightforward criteria were used would seem to suggest that this is a reasonable explanation. There is also the possibility that because calculus lends itself very easily to being embedded in a problem solving context, candidates are now rarely asked to perform routine calculus questions and thus do not have the opportunity to demonstrate their ability to carry out straightforward differentiation and integration as they did in the unrevised papers.

In geometry, the differences were the other way round ie results indicated more success in the revised papers. Previously, when straight forward instances of straight line geometry only were studied (reported under 'basic geometry') there were no significant differences. In the analysis above, the criteria included all instances of straight line geometry and all instances of criteria related to the geometry of the circle, to 3D geometry and to vectors. In the revised papers there were more of the difficult instances of relevant criteria and there were many more instances of criteria in total. For geometry, therefore, there appears to be a rise in standards across every band.

Presentation and the ability to work logically

During the marking process, markers were asked to grade candidates on their ability to set out work logically and to present their work clearly. Candidates were graded on a scale of 1 to 4 as an overall impression of their performance in these areas. The levels were described as follows:

1 Unsatisfactory

2 Weaknesses outweigh strengths

3 Strengths outweigh weaknesses

4 Major strengths

Table 5.9 shows the percentage of candidates at each band for their overall level of ability to set their work out logically.

Table 5.9: Percentage of candidates able to set out work logically

	Unrevised	Revised	Unrevised	Revised	Unrevised	Revised	Unrevised	Revised
Levels	Band A		Band B		Band C		Band D	
1	0	0	0	0	0	0	7	0
2	1	1	3	11	14	36	33	42
3	23	32	74	59	69	45	60	53
4	76	67	23	30	17	19	0	5

Table 5.10 shows the percentage of candidates at each band for their overall level of presentation.

Table 5.10: Percentage of candidates able to present work neatly

	Unrevised	Revised	Unrevised	Revised	Unrevised	Revised	Unrevised	Revised
Levels	Band A		Band B		Band C		Band D	
1	0	0	0	0	0	0	0	1
2	4	6	10	17	10	22	23	20
3	40	48	58	66	66	59	64	66
4	56	46	32	17	24	19	13	13

In general those candidates who were awarded higher bands set out their work more logically and provided a better presentation than candidates at the lower bands. There were minor differences between the unrevised and revised examinations. At band A greater percentage of candidates were assessed as showing major strengths in setting work out logically in the unrevised examination. At bands A and B a greater percentage of candidates were assessed as showing major strengths in presenting material clearly in the unrevised examination. These differences could be explained by the different demands of the examinations. In the unrevised examinations where the stages of problem solving are given, the need for clear presentation and logical working is restricted to a few lines of working. Candidates who can show major strengths in short problems may be unable to sustain this level of performance in more protracted problems as in the revised examination. However, none of these differences were statistically significant.

Summary

Contrasting unrevised and revised examinations

Several of the most important changes in the revised examination have no parallel in the unrevised examination. Some of these are illustrated earlier in this report and include such areas as working with unfamiliar mathematics and the embedding of routine knowledge and algorithms in real life contexts. Problem solving is another area where there can be no true comparison between the examinations. In the revised papers, candidates are expected to work their way through a problem consisting of several steps with little or no support. In the unrevised paper, the candidates are led though a problem step by step making it difficult for them to demonstrate their independent problem solving ability.

Discrete criteria

It was recognised from the start of this study that it would be difficult to make comparisons between two examinations which were so radically different in content and approach. In all the analyses carried out as part of the study, findings from consecutive years were combined to highlight the differences between performance in unrevised and revised examinations. The evidence appears to suggest that, for those individual criteria which were identified as being comparable across the unrevised and revised examinations, there is no consistent pattern of difference between candidates' performance and no statistically significant differences. For some criteria more candidates demonstrate attainment in the unrevised examination than in the revised examination. In other criteria, the trend is reversed. This is true of all bands from A to D.

Grouped criteria based on straightforward instances of criteria

Some of the criteria observed in the various examination papers studied were judged to be straightforward and others were judged to be more difficult. For one analysis, the straightforward criteria were grouped together under topic headings and a score was calculated for each candidate based on the number of correct responses divided by number of opportunities in each group. Candidates were then assigned to success or failure depending on a cut off score (see p90). The percentages of candidates who demonstrated attainment at each band were calculated and compared. The topic headings were algebra, geometry, trigonometry and calculus.

The figures for basic algebra and basic geometry showed different patterns of attainment. In algebra there were no differences in the percentages of successful candidates at bands A and B and minor differences at bands C and D where higher percentages of candidates showed attainment in the unrevised than in the revised examination. In straight line geometry again the differences were small but in this case a higher percentages of candidates showed attainment in the revised than in the unrevised examination. The same was true for both trigonometry and calculus except at band D in calculus where a larger percentage of candidates showed evidence of attainment in the unrevised higher examination. None of these differences were statistically significant.

Grouped criteria based on all instances of criteria

For this analysis all instances of criteria relating to each topic were included. The calculations were based on the total number of correct responses of all candidates at each band on five groups of criteria - algebra, geometry, trigonometry, differential calculus and integral calculus.

As a first step in the analysis, calculations were made on the attempt rate for each group of criteria and clear differences were noted between the two examinations. For the unrevised examination, rates tended to be relatively high across all bands except in trigonometry. In the revised examination, difference between the attempt rate for band A candidates and band D candidates ranged from 47% (algebra) to 20% (geometry). Differences of this kind are to be expected in view of the structure of the revised examination. Where questions are divided into discrete stages, almost all pupils can at least make an attempt at each stage. When the question is set as a problem to be solved and steps have to be determined by the candidates, lower attaining candidates will be unable to complete full questions and hence will be unable to demonstrate attainment of those criteria contained in the latter parts. This difference in question style reflects a fundamental and deliberate change in the revised mathematics syllabus. Successful candidates at Higher Grade mathematics are expected to solve problems rather than merely carry out routine procedures.

The second step involved calculating attainment on the basis of opportunities attempted. At bands A, B and C there were significant differences in integral calculus and at band B a significant difference was noted for differential calculus. These differences showed higher than statistically expected numbers of correct responses in the unrevised examinations. Geometry was the other area in which significant differences were noted. In this case, the results showed higher than expected numbers of correct responses in the revised examinations.

Conclusions

In considering the findings of this report, it must be remembered that only a relatively small number of candidates' papers were available for study and for this reason the following findings must be treated with caution.

- The changes to the syllabus and to the style of examination questions which include a focus on problem solving and mathematics in context mean that some opportunities provided in the examination are inaccessible to lower attaining candidates.
- There were no significant differences in candidates' ability to present their work clearly or set out their work in a logical manner.
- At the level of individual criteria, which were judged to be directly comparable between the unrevised and revised examinations, no significant differences were noted.
- When comparisons were made in relation to straightforward instances of criteria grouped under the headings of algebra, straight line geometry, trigonometry and calculus, there were no significant differences.
- When all instances of relevant criteria were included and comparisons were made between performances in the revised and unrevised examinations, there were significant differences at bands A, B and C in integral calculus and in bands B and D in differential calculus. In each case there were more correct responses in the unrevised examinations. Possible explanations for these differences are discussed above.
- There appears to be few opportunities for candidates in the revised examinations to demonstrate their ability to carry out routine procedures of differentiation and integration. Perhaps the mathematics panel for the Scottish Examination Board should consider if this is what is intended.
- When all instances of relevant criteria were included and comparisons were made between performances in the revised and unrevised examinations, there were significant differences at all bands in geometry. In each case there were more correct responses in the revised examinations.

Appendix — Mathematics

M.1 List of all criteria used in the analysis of the scripts

M.2 Bar graphs showing comparisons between the performance of candidates in unrevised and revised Higher Grade examinations on individual criteria (These criteria were selected because they were attempted by more than 50% of band C candidates and they were judged to be roughly comparable)

Appendix M.1

Higher Mathematics
List of criteria

Criterion Code/title†	1987 P2	1988 P2	1993 P1	1993 P2	1994 P1	1994 P2	
	Unrevised		Revised				
1. Eliminate one variable from two equations in three unknowns	xx		x	x			*
2. Sketch graph of y = - f(x)			x		x		
3. Sketch graph of y = f(x) + k			x		x		
4. Interpret graph				x			
5. Find equation of related graph			x				
6. Find formula for f(g(x))		xx	xx		x		**
7. Know that if f,g are inverses, f(g(x)) = x			x		x		
8. In $y = ae^{kx}$ substitute values of x,y			xx				
9. Know to solve 2 exponential equations simultaneously			x				
10. Solve $a^x = k$ by logs or otherwise					x		
11. Determine relationships $y = ax^b$				x			
12. Establish value for a in $y = ax^b$				x			
13. Establish value for b in $y = ax^b$				x			
14. Substitute co-ordinates into equation of a log graph			xx				
15. Use $k \log x = \log x^k$			x				
16. Use log 1/x = - log x					x		
17. Use anti-logs in log equation			x				
18. Solve simple log equation					x		
19. Identify discriminant	x	x	x	x		x	*
20. Apply condition for nature of roots	x	x	x	x		x	*
21. Use common factor (algebraic)		xx					
22. Verify a factor	x						
23. Use Remainder Theorem to find factor				xx			
24. Know to use given factor/root to find coefficient			x			x	
25. Identify remaining quadratic in synthetic division	x		x	xx		x	**
26. Convert root to index			xx				
27. Know that squared term is always positive					x		
28. Square out a bracket	x	xx	x			xxxx	**
29. Perform product of 2 binomials				xx			
30. Complete a square					x		
31. Identify coords of TP of quadratic from completed square					x		
32. Change quotient to separate fractions		x					
33. Simplify indices in a quotient		xx					
34. Change denominator of fraction to -ve power			x		x		
35. Solve a simple equation				x	x	xx	
36. Solve a simple inequation			x				
37. Solve quadratic with no degree in one term				x		x	
38. Use common factor in trigonometric equation	x	xxx					
39. Solve equation with fractional indices				x			
40. Substitute linear equation in quadratic	x	x			x	x	**
41. Simplify to quadratic	x	x			xx	x	**
42. Factorise quadratic	8	xxxx	xxxx	6	xx	xxxx	**
43. Identify roots from factors	10	6	6	6	xxxx	xxxx	**
44. Identify recurrence factor				x			
45. Appreciate notion of limit in recurrence situation				x	x		
46. Given recurrence formula find u_2					x		
47. Find smallest n for given value of u_n					x		
48. Given formula for S_n find u_1 and u_2	x						
49. Given 2 terms of AP find common difference	x						
50. Know formula for sum of AP	x						
51. Substitute correctly in formula	x					x	*

52. Know condition for geometric series		x					
53. Know condition for sum to infinity		x					
54. Find image points under matrix transformation	xx	xxxxx					
55. Recognise a dilatation	x						
56. Find equation of image line in matrix transformation	x						
57. Extended processing of matrices	x						
58. Find area of triangle by 1/2 bh				x		xxx	
59. Use sum of angles of a triangle			x				
60. Know radius and tangent are perpendicular	x			x	x		**
61. Use distance formula with coordinates of 2 points	x	xx		x	x	xx	**
62. Find mid-point of line joining 2 points	x	xx			x		**
63. Substitute coordinates in gradient formula	xxx	xx	x		xx	x	**
64. Know that gradient is tangent of angle with x axis			x		x	x	
65. Identify gradient in equation of line			xx	x		xx	
66. Use gradients of perpendicular lines result	xx		x	x	x	xx	**
67. Find equation of line by inspection		xx					
68. Substitute point and gradient in equation of line	x		x	x		xx	**
69. Find equation of line in terms of log x, log y				x			
70. Find coordinates of point on a line		xxxx					
71. Show that points lie on a line			x				
72. Use coord of point to find unknown coefficient in equation						xx	
73. Find intersection of line and y axis	x						
74. Know to solve simultaneous eqn to find point of intersection				xx			
75. Find intersection of two lines	x			x		x	**
76. Find intersection of curve and y axis						x	
77. Find intersection of curve and x axis				xx			
78. Know how to find intersection of curve and y=k						x	
79. Relate roots to intercepts on x axis		x				xxx	*
80. Verify that point lies on a circle		xx					
81. From $(x-a)^2 + (y-b)^2 = r^2$ find radius			x				
82. From $(x-a)^2 + (y-b)^2 = r^2$ find centre			x		x		
83. Use centre, radius to give equation of circle	x	x	x	x	x		**
84. From general equation of circle, identify radius			x			xx	
85. From general equation of circle, identify centre						xx	
86. Relate equal roots to tangency	x			x			*
87. From equal roots find coords of point of contact	x						
88. Identify coordinates in 3D diagram				xxx	x	xxx	
89. Use distance formula with 3D coords					x	x	
90. Find midpoint in 3D coordinates				x		x	
91. Find 3D coordinates of point dividing line internally	x						
92. Find 3D coordinates of point dividing line externally		x					
93. Find 3D coordinates of image of point in line		x					
94. Find 3D vector components from coordinates	xx	xx	x		xxx	xx	**
95. Set up vector equation					x		
96. Solve vector equation to find 3D coordinates of point					x		
97. Find 2D vector components from coordinates	xx						
98. Know how to calculate vector length			x	xx			
99. Know condition for collinearity		x			x		*
100. Interpret unit vector as column vector			xxx				
101. Find sum or difference of column vectors					xx		
102. Evaluate scalar product of column vectors	x	x	xx		x	x	**
103. Know how to find angle between vectors a b cos ϑ	x					x	**
104. Know that scalar product =0, implies perpendicular		x	x		x		**
105. Express directed line as difference of position vectors		xxxxx					
106. Form equation with position vectors		xx					
107. Know exact value of sin, cos, tan					xxxx		
108. Know the effect of c on the graph sincx/coscx					x		
109. Know the effect of b on the graph bsinx/bcosx					x		

110. Know the effect of a on the graph a+sinx/a+cosx					x		
111. Given sin, cos, tan find 1st quad angle (degrees)	6	xx	xx	xx	xx	xxxxx	**
112. Given 1st quadrant angle (degrees) find related angle	xxx	xx	x		x	x	**
113. Given sin, cos, tan find angle (radians)			xx	x	x		
114. Given 1st quadrant angle (radians) find related angle				x	x		
115. Use trig ratio in right angled triangles						x	
116. Solve simple trig equation to find value of trig ratio				x	x		
117. Factorise trig quadratic			xx		xx		
118. Use $\sin^2 + \cos^2 = 1$ in substitution			xx				
119. Use expansion of sin2x, cos2x	x	xx			xxx		**
120. Use 2cosxsinx = sin2x	x		x				
121. Use formulae for sum, difference of sin and cos	xx	x					
122. Use formulae for sin, cos, (A+B), (A-B)	x				xx		*
123. Use formula for $\cos^2$ or $\sin^2$	x	xx					
124. Given exact sin or cos find other			xx		xx		
125. Substitute in given trig formulae			x				
126. Find area of triangle by 1/2 absin C				xx	x		
127. Use acosx + bsinx = rcos(x-c) or similar	x			x		x	**
128. Know how to find r in rcos(x-c)	x			x		x	**
129. Know how to find angle c in rcos(x-c)	x			x		x	**
130. Know to use rcos(x-c) or similar to solve equation	x			x		x	**
131. Know to use rcos(x-c) or similar to solve inequation	x					x	
132. Solve equation of form rcos(x-c) = k		x		x		x	**
133. Solve inequation of form rcos(x-c) ≤ k						x	
134. Find max or min value of rcos(x-c)		xx					
135. Find x where max/min value of rcos(x-c) occurs		xx					
136. Differentiate a polynomial	x	xx	x	x	xx	xxx	**
137. Differentiate +ve fractional index		x	x				**
138. Differentiate -ve index		x	x		x		*
139. Differentiate sinax/cosax (multiple angles)					x		
140. Differentiate using chain rule (algebra)				x			
141. Use chain rule in trig differentiation for multiple angles	x		x				**
142. Use chain rule in trig differentiation for power	x				x		**
143. Know to differentiate for gradient	x				x	xx	**
144. Subst x coord to get gradient at a point					x	x	
145. Equate derivative to gradient to find x coordinate	x					x	*
146. Know to differentiate to find stationary points	x	x	x	x			**
147. Find y values of roots of derivative/evaluate function given x	2	6		4	3	11	**
148. Use table or 2nd derivative for nature of stationary points	xx	xx		xxx		x	**
149. Equate dy/dx to zero for stationary points	x	xx	x	xx		x	**
150. Know to differentiate for greatest/least values		x		x		x	*
151. Know to differentiate to determine inc/dec nature			x				
152. Know that f'(x) < 0 for decreasing			x				
153. Sketch graph to show stat points, intersect on axes	x	xx					
154. Know to integrate f'(x) to find f(x)	x		x				**
155. Integrate a constant			x				
156. Integrate a polynomial	x		x	x	x	x	*
157. Integrate term with -ve power	x						
158. Integrate term with fractional index	x						
159. Introduce constant of integration	x		xx		xx		**
160. Find constant of integration by subst.	x		x				**
161. Deal with integration of power $(px+q)^n$, n +ve integer		x					
162. Deal with integration of power $(px+q)^n$, n fractional			x				
163. Deal with coefficient of x in integration of $(px+q)^n$		x	x				**
164. Integrate sinax, cosax (with multiple angle)			x				
165. Integrate by anti-differentiation					x		
166. Identify (by limits) area under curve	x			x		x	*
167. Substitute limits in definite integral	x	x	x	x		x	**

168. Know features of graph of $f'(x)$ from graph of $f(x)$			x				
169. Appreciate requirements of proof	xx	xx		x			*
170. Introduce statement of limit	x	x					
171. Demonstrate ability in unfamiliar situations	x	xx		xxxx			*
172. Implement a given procedure						xx	
173. Perform algebraic simplification	2	9	2	5	1	7	**
174. Perform necessary exponential calculations			x				
175. Perform necessary arithmetic with fractional indices			xx				
176. Perform necessary arithmetic	8	12	9	7	8	14	*
177. Carry out continuous data processing		xx		xxx		xx	*
178. Marshal facts to show given result		x		xx		x	*
179. Recognise a pattern	x	x				x	*
180 Show acceptable precision in language	x	xxx	xxxx	xxx	xx	x	*

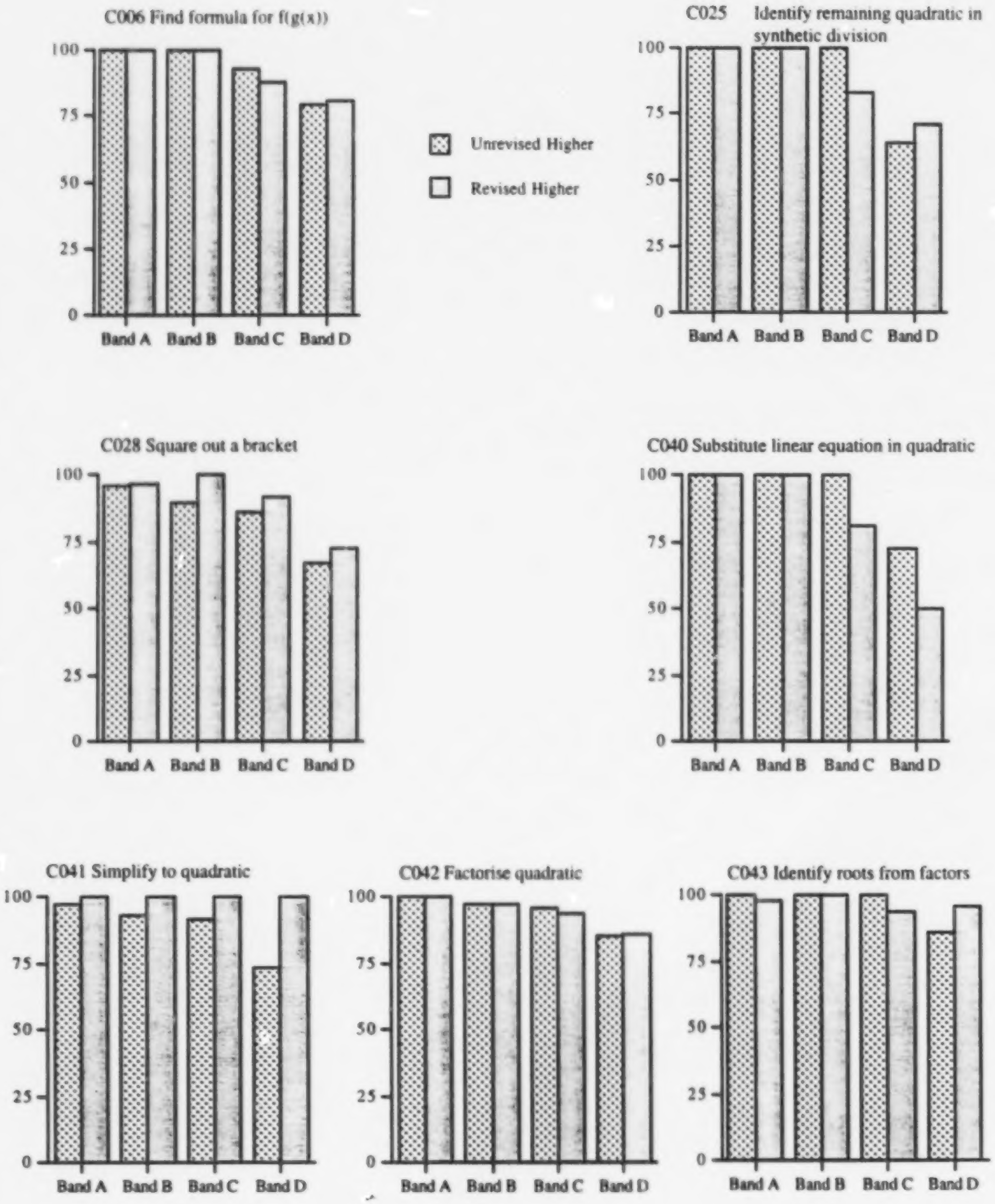
† Note that for reasons of space, the criteria are written in an abbreviated form

* Criteria which appear in both the unrevised and the revised examinations of the selected years

** Criteria which were judged to be comparable

Appendix M.2

Algebra



C006 Find formula for f(g(x))

C025 Identify remaining quadratic in synthetic division

C028 Square out a bracket

C040 Substitute linear equation in quadratic

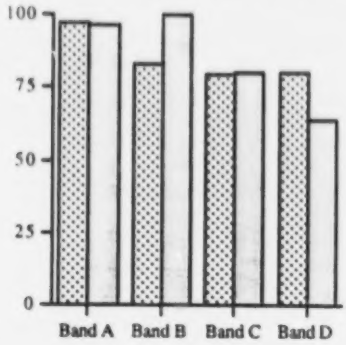
C041 Simplify to quadratic

C042 Factorise quadratic

C043 Identify roots from factors

Geometry

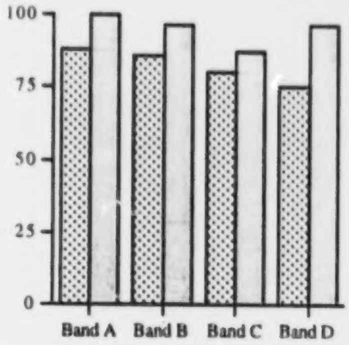
C060 Know radius and tangent are perpendicular



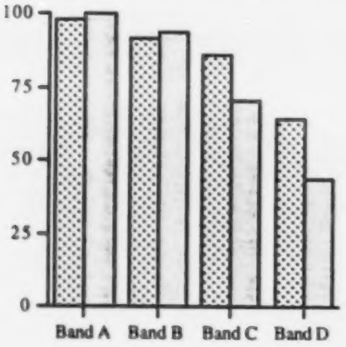
▨ Unrevised Higher

☐ Revised Higher

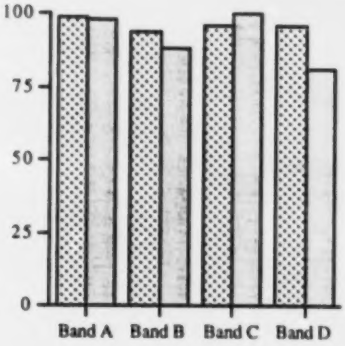
C061 Use distance formula with co-ords of 2 points



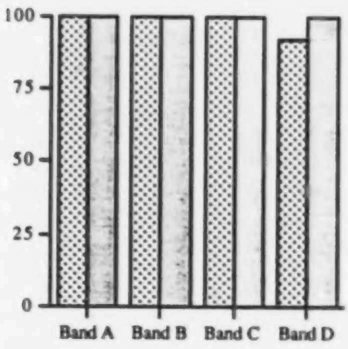
C062 Find mid-point of line joining 2 points



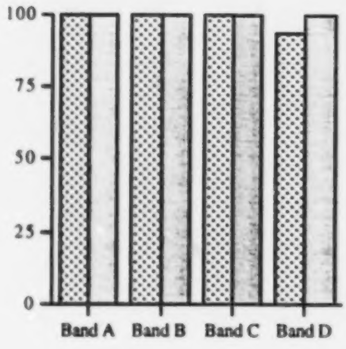
C063 Subst co-ords in gradient formula



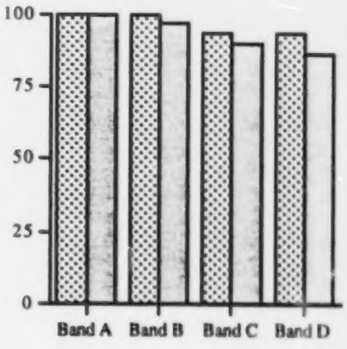
C066 Use result from gradient of perpendicular lines



C068 Subst point and gradient in equn of line

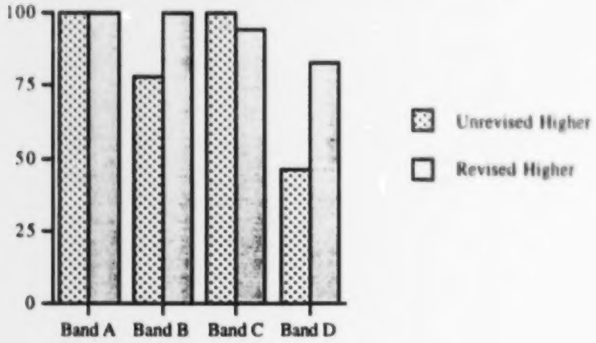


C075 Find intersection of two lines

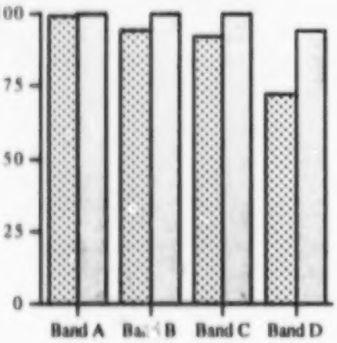


Geometry/Vectors

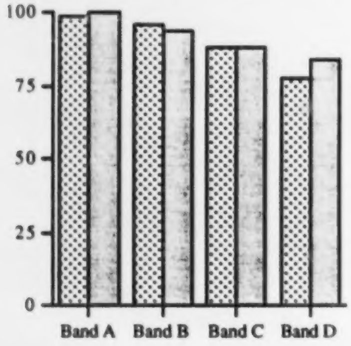
C083 Use centre, radius to give equn of circle



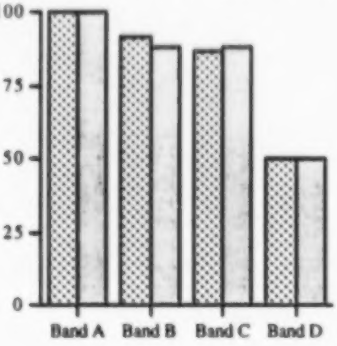
C094 Find 3D vector components from co-ords



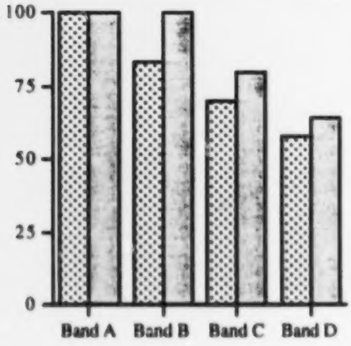
C102 Evaluate scalar product of column vectors



C103 Know how to find the angle between vectors

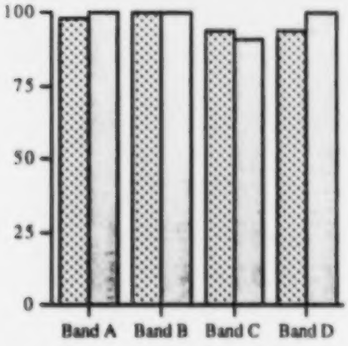


C104 Know that scalar product=0, implies perpendicular



Trigonometry

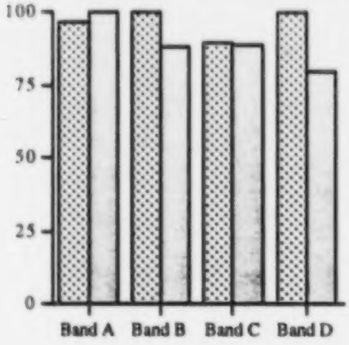
C111 Given trig ratio, find angle



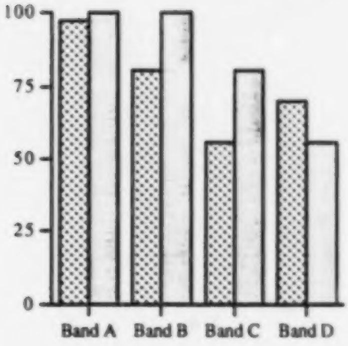
▨ Unrevised Higher

☐ Revised Higher

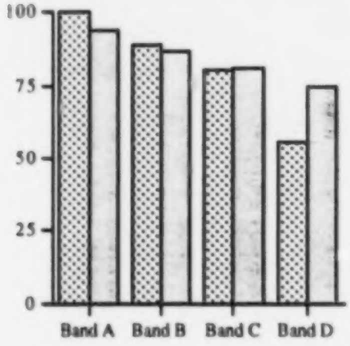
C112 Given angle find related angle



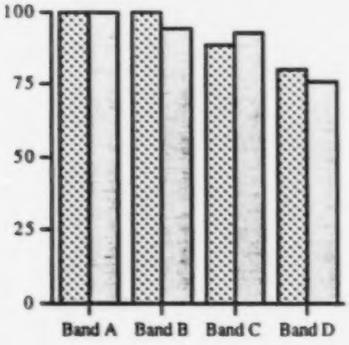
C119 Use expansion of sin2x etc.



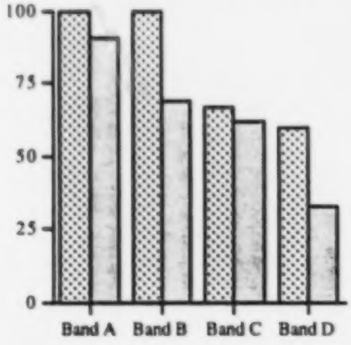
C127 Use acosx + bsinx = rcos(x-c)



C128 Know how to find r in rcos(x-c)



C129 Know how to find angle C in rcos(x-c)

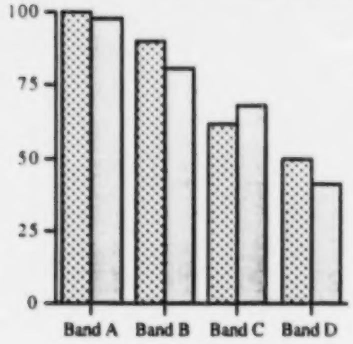


Trigonometry/Calculus (Differential)

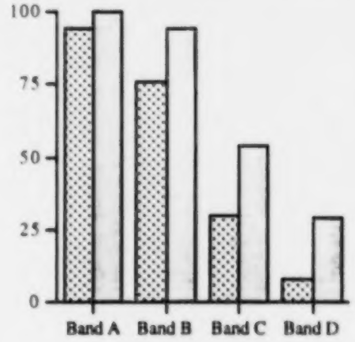
▨ Unrevised Higher

☐ Revised Higher

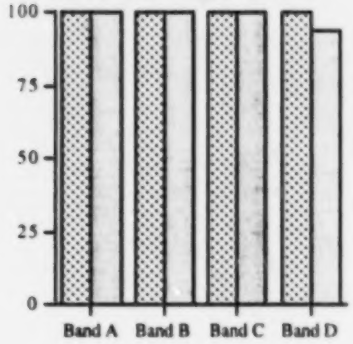
C130 Know to use rcos(x-c) to solve equations



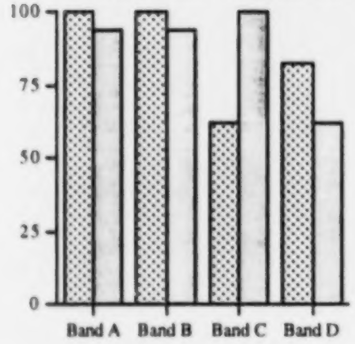
C132 Solve equation of form rcos(x-c)



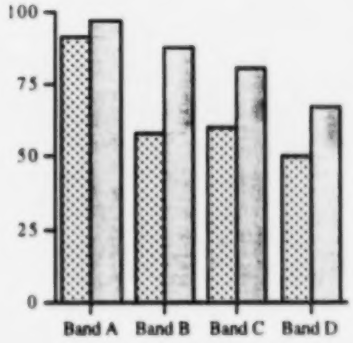
C136 Differentiate a polynomial



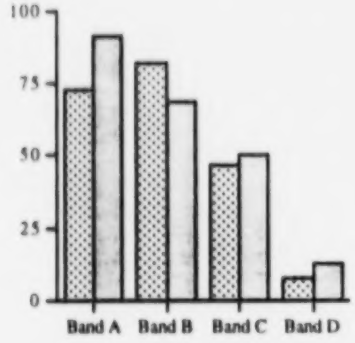
C137 Differentiate +ve fractional index



C141 Use chain rule in trig differentiation for coefficient

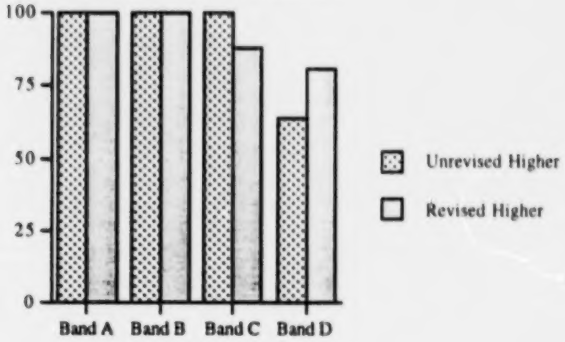


C142 Use chain rule in trig differentiation for power

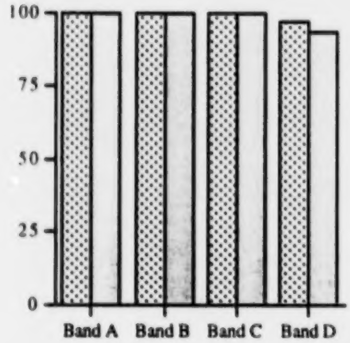


Calculus (Differential)

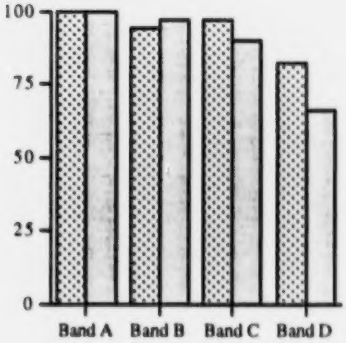
C143 Know to differentiate for gradient



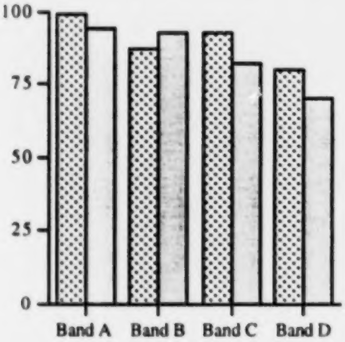
C146 Know to differentiate to find stationary points



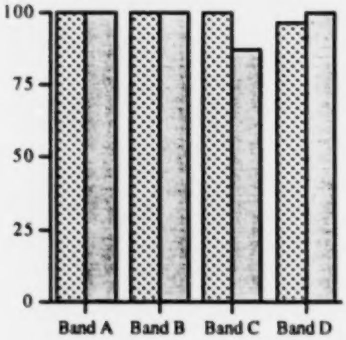
C147 Find y values of roots



C148 Use table or 2nd derivative for nature of stationary points

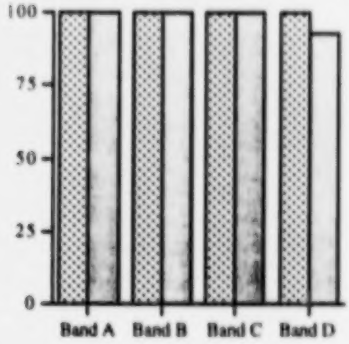


C149 Equate dy/dx to zero for stationary points



Calculus (Integral)

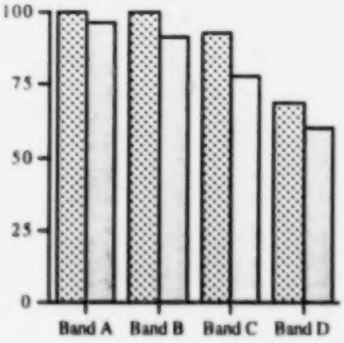
C154 Know to integrate f'(x) to find f(x)



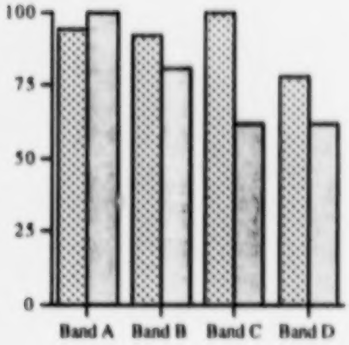
▨ Unrevised Higher

☐ Revised Higher

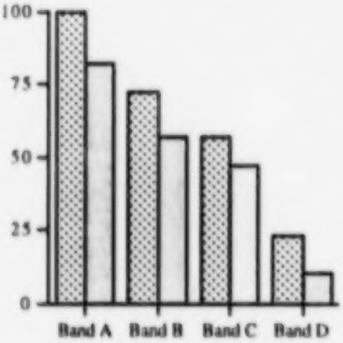
C159 Introduce constant of integration



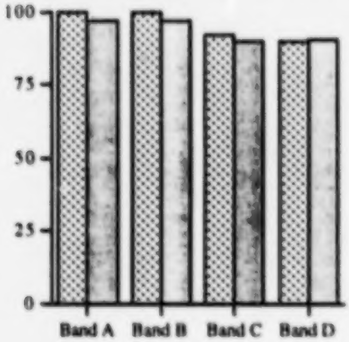
C160 Find constant of integration by subs



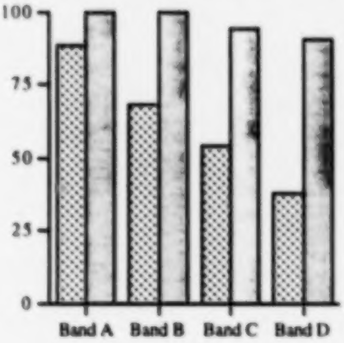
C163 Deal with coeff of x in integration of $(px + q)^n$



C167 Substitute limits in definite integral



Basic Algebra

C173 Perform algebraic simplification